An Introduction to Geometric Topology

Bruno Martelli

Bruno Martelli
University of Pisa, Italy

`http://people.dm.unipi.it/martelli/`

To Ada, Sara, and Ylenia.

Contents

Introduction

The aim of this book is to introduce the reader to an area of mathematics called *geometric topology*. The text should be suitable to a master or PhD student in mathematics interested in geometry, and more generally to any curious mathematician with a standard background in topology and analysis.

We interpret the term "geometric topology" in a quite restrictive sense: for us, this topic is the study of topological manifolds via the assignment of a preferred "geometric structure," that is some Riemannian metric of a particular nice kind. It is one of the most important discoveries in modern geometry that such preferred metrics exist on every compact manifold of dimension two and three, and this is exactly what this book is about.

In other words, this book is an introduction to surfaces and three-manifolds, and to their *geometrisation*, due to Poincaré and Koebe in 1907 in dimension two and to Thurston and Perelmann in 2002 in dimension three. Therefore this is also a textbook on *low-dimensional topology*, except that we completely neglect four-manifolds, that form a relevant part of this area but which do not (yet?) fit in any geometrisation perspective.

There are already many good textbooks on surfaces, so our main new contribution is probably to furnish a complete introduction to Thurston's geometrisation of three-manifolds, that includes both the topological side of the story (the decomposition of three-manifolds along spheres and tori, the classification of Seifert manifolds) and the geometric side (hyperbolic geometry, Thurston's Dehn filling Theorem, Mostow rigidity, the eight geometries).

This book is essentially self-contained: in the first chapter we recall all the relevant background material in differential topology and geometry, and from the second chapter on we never mention a theorem without giving a proof – with only one important but unavoidable exception: Perelman's solution of the geometrisation conjecture, stated in Chapter 12.

Outline. Here is a brief outline of the material contained in each chapter. The book is divided into three parts. In the first, we introduce hyperbolic geometry and its relatives, the elliptic and flat geometries. Hyperbolic geometry is by far the richest, the most beautiful, and also the most important geometry in dimensions two and three: its relevance is witnessed

by the folk sentence that a "generic" manifold of dimension two or three has a hyperbolic structure, that is a Riemannian structure locally isometric to the hyperbolic plane or space.

Chapter 1 includes all the preliminaries that we will need: in particular, we quickly review various general notions of differential topology and geometry that can be found in many excellent books.

Chapter 2 introduces the reader to the hyperbolic space $\mathbb{H}^n$: as opposite to the sphere S^n and to the euclidean space $\mathbb{R}^n$, the hyperbolic space may be represented in at least four different ways, none of which is prevalent: each representation is called a *model*. We introduce the models, and then study the isometries, the compactification of $\mathbb{H}^n$, and its subspaces.

Chapter 3 deals with hyperbolic manifolds. Maybe unexpectedly, these objects are related to a number of different beautiful mathematical concepts like discrete subgroups of Lie groups, polyhedra, and tessellations. We also present some generalisations like manifolds with geodesic boundary, cone manifolds, and orbifolds.

Chapter 4 introduces the *thick-thin decomposition*, a general structural theorem on hyperbolic manifolds that allows us to understand in particular the topology of the complete hyperbolic manifolds that have finite volume but are not compact. It also contains various information on flat and elliptic manifolds, including Bieberbach's Theorem.

Chapter 5 introduces a variety of notions and results on hyperbolic manifolds that are connected with the points at infinity of $\mathbb{H}^n$.

The second part of the book deals with surfaces. We classify and geometrise every surface of finite type in Chapter 6, where we also study simple closed curves in surfaces and define the *mapping class group*.

Chapter 7 presents the *Teichmüller space* of a genus-g surface, as the space of its hyperbolic structures. We use the Fenchel–Nielsen coordinates to show that this space is in fact homeomorphic to $\mathbb{R}^{6g-6}$.

Chapter 8 introduces Thurston's beautiful theory of diffeomorphisms of surfaces. We introduce Thurston's compactification of the Teichmüller space, and the consequent classification of the elements of the mapping class group into three classes. We define and study some intriguing objects called *geodesic currents* and *laminations*.

The last (and longest) part of the book is devoted to three-manifolds. In Chapter 9 we introduce the first topological facts, including the prime decomposition, incompressible surfaces, and Haken manifolds.

Chapter 10 is entirely devoted to Seifert manifolds, a class of three-manifolds that contains many interesting examples. We classify these manifolds completely.

In Chapter 11 we present various techniques that topologists use every day to construct three-manifolds: Heegaard splittings, knots and links, Dehn surgery, and surface bundles. We end by stating and proving the canonical torus decomposition.

In Chapter 12 we finally move from topology to geometry: there are eight relevant geometries in dimension three, and we introduce them with some detail. We show in particular that the Seifert manifolds occupy precisely six of them.

In Chapter 13 it is due time to start investigating the most interesting of the three-dimensional geometries, the hyperbolic one: we prove that in dimension three every manifold has at most one hyperbolic structure, and this important fact is known as *Mostow's rigidity Theorem*.

In Chapter 14 we construct many examples of hyperbolic 3-manifolds, by introducing ideal triangulations and *Thurston's equations*.

Chapter 15 contains a complete proof of Thurston's *hyperbolic Dehn filling theorem* and a discussion on the volumes of hyperbolic 3-manifolds.

As we mentioned above, this book is almost entirely self-contained and the bibliography is minimised to the strict necessary: each chapter ends with a short section containing the pertaining references, that consist essentially in the sources that we have consulted for that chapter. Many of the topics presented here have their origin in Thurston's notes and papers and are of course already covered by other books, that we have widely used, so our bibliography consists mainly of secondary sources.

- Figure 2.4 (tessellation of hyperbolic plane), Figure 2.13 (horo-sphere), Figure 4.2 (pseudosphere), and Figure 3.6 (tessellations of hyperbolic plane) created by Claudio Rocchini,
- Figure 3.3-right (dodecahedron) created by DTR,
- Figure 3.9 and 3.10 (dodecahedral tessellations) created by Roice3,
- Figure 3.14 (Voronoi tessellation) created by Mysid and Cyp,
- Figure 3.15 (fundamental domain of the modular group) and Figure 11.10 (Seifert surface for the trefoil knot) created by Kilom691,
- Figure 6.10 created by Tttrung,
- Figure 8.4 created by Adam Majewski,
- Figure 10.9 (The Hopf fibration) created by Niles Johnson,
- Figure 11.3 (Knot sum) originally created by Maksim and tP-brocks13,
- Figure 11.9 (Satellite knots) originally created by RyBu.

The painting in the front page is a courtesy of Mariette Michelle Egreteau. The Figures 5.1, 5.2, 5.3 and 5.4 were made by Jacopo Guoyi Chen.

Part 1

Hyperbolic geometry

Preliminaries

We expect the reader to be familiar with the mathematics usually taught to undergraduates, like multivariable differential calculus, group theory, topological spaces, and fundamental groups.

Some knowledge on differentiable and Riemannian manifolds would also help, at least on an intuitive way: in any case, this chapter introduces from scratch everything we need from differential topology and geometry. We also include some important information on groups (like Lie groups and group actions), and a few basic notions of measure theory and algebraic topology, with a very quick overview of homology theory. Finally, we use these tools to introduce cell complexes, handle decompositions, and triangulations: these are the main practical instruments that we have to build manifolds concretely.

Most results stated here are given without a proof: details can be found in various excellent books, some of which will be cited below.

1.1. Differential topology

We introduce manifolds, bundles, embeddings, tubular neighbourhoods, isotopies, and connected sums. Throughout this book, we will always work in the smooth category. The material contained in this section is carefully explained in *Differentiable manifolds* of Kosinksi [**34**].

1.1.1. Differentiable manifolds. A *topological manifold* of dimension n is a paracompact Hausdorff topological space M locally homeomorphic to $\mathbb{R}^n$. In other words, there is a covering $\{U_i\}$ of M consisting of open sets U_i homeomorphic to open sets V_i in $\mathbb{R}^n$.

Topological manifolds are difficult to investigate, their definition is too general and allows to directly define and prove only few things. Even the notion of dimension is non-trivial: to prove that an open set of $\mathbb{R}^k$ is not homeomorphic to an open set of $\mathbb{R}^h$ for different k and h we need to use non-trivial constructions like homology. It is also difficult to treat topological subspaces: for instance, the *Alexander horned sphere* shown in Figure 1.1 is a subspace of $\mathbb{R}^3$ topologically homeomorphic to a 2-sphere. It is a complicated object that has many points that are not "smooth" and that cannot be "smoothened" in any reasonable way.
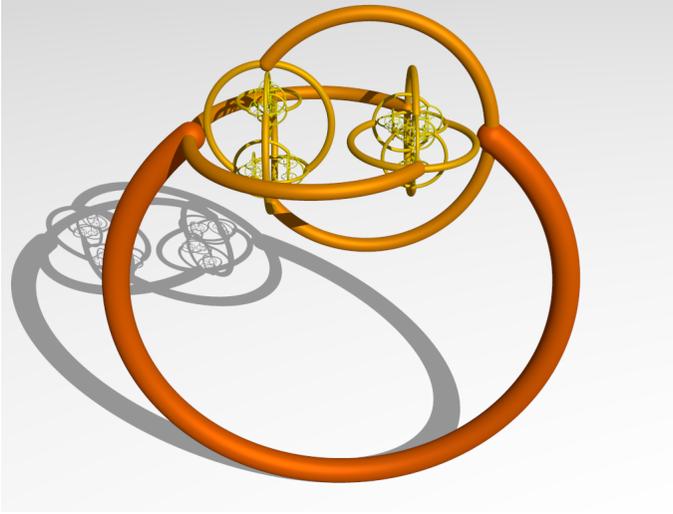
Figure 1.1. The *Alexander horned sphere* is a subset of $\mathbb{R}^3$
homeomorphic to the 2-sphere S^2. It divides $\mathbb{R}^3$ into two
connected components, none of which is homeomorphic to
an open ball. It was constructed by Alexander as a coun-
terexample to a natural three-dimensional generalisation of
Jordan's curve theorem. The natural generalisation would
be the following: does every 2-sphere in $\mathbb{R}^3$ bound a ball?
If the 2-sphere is only topological, the answer is negative
as this counterexample shows. If the sphere is a differen-
tiable submanifold, the answer is however positive as proved
by Alexander himself.

We need to define some "smoother" objects, and for that purpose
we can luckily invoke the powerful multivariable infinitesimal calculus. Let
$U \subset \mathbb{R}^n$ be an open set: a map $f \colon U \to \mathbb{R}^k$ is *smooth* if it is C^∞, *i.e.* it has
partial derivatives of any order.

Definition 1.1.1. Let M be a topological manifold. A *chart* is a fixed
homeomorphism $\varphi_i \colon U_i \to V_i$ between an open set U_i of M and an open
set V_i of $\mathbb{R}^n$. An *atlas* is a set of charts $\{(U_i, \varphi_i)\}$ where the open sets U_i
cover M.

If $U_i \cap U_j \neq \varnothing$ there is a *transition map* $\varphi_{ji} = \varphi_j \circ \varphi_i^{-1}$ that sends
homeomorphically the open set $\varphi_i(U_i \cap U_j)$ onto the open set $\varphi_j(U_i \cap U_j)$.
Since these two open sets are in $\mathbb{R}^n$, it makes sense to require φ_{ij} to be
smooth. The atlas is *differentiable* if all transition maps are smooth.

Definition 1.1.2. A *differentiable manifold* is a topological manifold
that is equipped with a differentiable atlas.
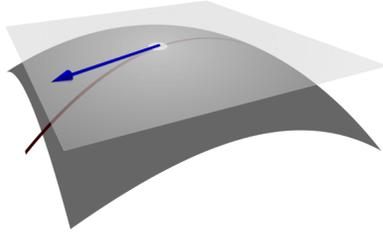
Figure 1.2. The tangent space at p is defined as the set of all curves γ with $\gamma(0) = p$ seen up to an equivalent relation that identifies two curves having (in some chart) the same tangent vector at p. This is a chart-independent condition.

We will often use the word *manifold* to indicate a differentiable manifold. The integer n is the *dimension* of the manifold. We have defined the objects, so we now turn to their morphisms.

Definition 1.1.3. A map $f\colon M \to M'$ between differentiable manifolds is *smooth* if it is smooth when read locally through charts. This means that for every $p \in M$ there are two charts (U_i, φ_i) of M and (U'_j, φ'_j) of N with $p \in U_i$ and $f(p) \in U'_j$ such that the composition $\varphi'_j \circ f \circ \varphi_i^{-1}$ is smooth wherever it is defined.

A *diffeomorphism* is a smooth map $f\colon M \to M'$ that admits a smooth inverse $g\colon M' \to M$. A *curve* in M is a smooth map $\gamma\colon I \to M$ defined on some open interval I of the real line, which may be bounded or unbounded.

Definition 1.1.4. A differentiable manifold is *oriented* if it is equipped with an atlas where all transition functions are orientation-preserving (that is, the determinant of their differential at any point is positive).

A manifold which can be oriented is called *orientable*.

1.1.2. Tangent space. Let M be a differentiable manifold of dimension n. We may define for every point $p \in M$ a n-dimensional vector space $T_p M$ called the *tangent space*.

The space $T_p M$ may be defined as the set of all curves $\gamma\colon (-a, a) \to M$ such that $\gamma(0) = p$ and $a > 0$ is arbitrary, considered up to some equivalence relation. The relation is the following: we identify two curves that, read on some chart (U_i, φ_i), have the same tangent vector at $\varphi_i(p)$. The definition does not depend on the chart chosen.

A chart identifies $T_p M$ with the usual tangent space $\mathbb{R}^n$ at $\varphi_i(p)$ in the open set $V_i = \varphi_i(U_i)$. Two distinct charts φ_i and φ_j provide different identifications with $\mathbb{R}^n$, which differ by a linear isomorphism: the differential

$d\varphi_{ji}$ of the transition map φ_{ij}. The structure of T_pM as a vector space is then well-defined, while its identification with $\mathbb{R}^n$ is not.

Every smooth map $f\colon M \to N$ between differentiable manifolds induces at each point $p \in M$ a linear map $df_p\colon T_pM \to T_{f(p)}N$ between tangent spaces in the simplest way: the curve γ is sent to the curve $f \circ \gamma$.

Definition 1.1.5. A smooth map $f\colon M \to N$ is a *local diffeomorphism* at a point $p \in M$ if there are two open sets $U \subset M$ and $V \subset N$ containing respectively p and $f(p)$ such that $f|_U\colon U \to V$ is a diffeomorphism.

The inverse function theorem in $\mathbb{R}^n$ implies easily the following fact, that demonstrates the importance of tangent spaces.

Theorem 1.1.6. *Let $f\colon M \to N$ be a smooth map between manifolds of the same dimension. The map is a local diffeomorphism at $p \in M$ if and only if the differential $df_p\colon T_pM \to T_{f(p)}N$ is invertible.*

In this theorem a condition satisfied at a single point (differential invertible at p) implies a local property (local diffeomorphism). Later on, we will see that in Riemannian geometry a condition satisfied at a single point may also imply a global property.

If $\gamma\colon I \to M$ is a curve, its *velocity* $\gamma'(t)$ in $t \in I$ is the tangent vector $\gamma'(t) = d\gamma_t(1)$. Here "1" means the vector 1 in the tangent space $T_tI = \mathbb{R}$. We note that the velocity is a vector and not a number: the modulus of a tangent vector is not defined in a differentiable manifold (because the tangent space is just a real vector space, without a norm).

1.1.3. Differentiable submanifolds. Let N be a differentiable manifold of dimension n.

Definition 1.1.7. A subset $M \subset N$ is a *differentiable submanifold* of dimension $m \leqslant n$ if every $p \in M$ has an open neighbourhood $U \subset N$ diffeomorphic to an open set $V \subset \mathbb{R}^n$ via a map $\varphi\colon U \to V$ that sends $U \cap M$ onto $V \cap L$, where L is a linear subspace of dimension m.

The pairs $\{(U \cap M, \varphi|_{U \cap M})\}$ form an atlas for M, which inherits a structure of m-dimensional differentiable manifold. At every point $p \in M$ the tangent space T_pM is a linear subspace of T_pN.

1.1.4. Fibre bundles. We introduce a particularly nice kind of maps.

Definition 1.1.8. A *smooth fibre bundle* is a smooth map

$$\pi\colon E \longrightarrow B$$

such that every fibre $\pi^{-1}(p)$ is diffeomorphic to a fixed manifold F and π looks locally like a projection. This means that B is covered by open sets U_i equipped with diffeomorphisms $\psi_i\colon U_i \times F \to \pi^{-1}(U_i)$ such that $\pi \circ \psi_i$ is the projection on the first factor.

The manifolds E and B are called the *total* and *base manifold*, respectively. The manifold F is the *fibre* of the bundle. A *section* of the bundle is a smooth map $s\colon B \to E$ such that $\pi \circ s = \mathrm{id}_B$. Two fibre bundles $\pi\colon E \to B$ and $\pi'\colon E' \to B$ are *isomorphic* if there is a diffeomorphism $\psi\colon E \to E'$ such that $\pi = \pi' \circ \psi$.

1.1.5. Vector bundles. A *smooth vector bundle* is a smooth fibre bundle where every fibre $\pi^{-1}(p)$ has the structure of a *n*-dimensional real vector space which varies smoothly with p. Formally, we require that $F = \mathbb{R}^n$ and $\psi_i(p, \cdot)\colon F \to \pi^{-1}(p)$ be an isomorphism of vector spaces for every ψ_i as above.

The *zero-section* of a smooth vector bundle is the section $s\colon B \to E$ that sends p to $s(p) = 0$, the zero in the vector space $\pi^{-1}(p)$. The image $s(B)$ of the zero-section is typically identified with B via s.

Two vector bundles are *isomorphic* if there is a diffeomorphism ψ as above, which restricts to an isomorphism of vector spaces on each fibre. As every manifold here is differentiable, likewise every bundle will be smooth and we will hence often omit this word.

1.1.6. Tangent and normal bundle. Let M be a differentiable manifold of dimension n. The union of all tangent spaces

$$TM = \bigcup_{p \in M} T_p M$$

is naturally a differentiable manifold of double dimension $2n$, called the *tangent bundle*. The tangent bundle TM is naturally a vector bundle over M, the fibre over $p \in M$ being the tangent space $T_p M$.

Let $M \subset N$ be a smooth submanifold of N. The *normal space* at a point $p \in M$ is the quotient vector space $\nu_p M = T_p N / T_{p M}$. The *normal bundle* νM is the union

$$\nu M = \bigcup_{p \in M} \nu_p M$$

and is also naturally a smooth vector bundle over M. The normal bundle is not canonically contained in TN like the tangent bundle, but (even more usefully) it may be embedded directly in N, as we will soon see.

1.1.7. Vector fields. A *vector field* X on a smooth manifold M is a section of its tangent bundle. A point p determines an *integral curve* $\alpha\colon I \to M$ starting from p, that is a smooth curve with $\alpha(0) = p$ and $\alpha'(t) = X(\alpha(t))$ for all $t \in I$. The curve α is unique if we require the interval I to be maximal. It depends smoothly on p.

1.1.8. Immersions and embeddings. A smooth map $f\colon M \to N$ between manifolds is an *immersion* if its differential is everywhere injective: note that this does not imply that f is injective. The map is an *embedding* if it is a diffeomorphism onto its image: this means that f is injective, its image is a submanifold, and $f\colon M \to f(M)$ is a diffeomorphism.

Theorem 1.1.9. *If M is compact, every injective immersion f : M → N is an embedding.*

1.1.9. Isotopy and ambient isotopy. Let X and Y be two topological spaces. We recall that a *homotopy* between two continuous maps $\varphi, \psi \colon X \to Y$ is a continuous map $F \colon X \times [0, 1] \to Y$ such that $F_0 = \varphi$ and $F_1 = \psi$, where $F_t = F(\cdot, t)$.

Let M and N be differentiable manifolds. A *smooth isotopy* between two embeddings $\varphi, \psi \colon M \to N$ is a smooth homotopy F between them, such that every map F_t is an embedding. Again, we will shortly use the word *isotopy* to mean a smooth isotopy. We note that many authors do not require an isotopy to be smooth, but we do.

Being isotopic is an equivalence relation for smooth maps $M \to N$: two isotopies F_t and G_u can be glued if $F_1 = G_0$, and since we want a smooth map we priorly modify F and G so that they are constant near $t = 1$ and $u = 0$. This can be done easily by reparametrising t and u.

An *ambient isotopy* on N is an isotopy between id_N and some other diffeomorphism $\varphi \colon N \to N$, such that every level is a diffeomorphism. Two embeddings $\varphi, \psi \colon M \to N$ are *ambiently isotopic* if there is an ambient isotopy F of N such that $\psi = F_1 \circ \varphi$.

Theorem 1.1.10. *If M is compact, two embeddings $\varphi, \psi \colon M \to N$ are isotopic if and only if they are ambiently isotopic.*

1.1.10. Tubular neighbourhood. Let $M \subset N$ be a differentiable submanifold. A *tubular neighbourhood* of M is an open subset $U \subset N$ with a diffeomorphism $\nu M \to U$ sending identically the zero-section onto M.

Theorem 1.1.11. *Let $M \subset N$ be a closed differentiable submanifold. A tubular neighbourhood for M exists and is unique up to an isotopy fixing M and up to pre-composing with a bundle isomorphism of νM.*

If we are only interested in the open set U and not its parametrisation, we can of course disregard the bundle isomorphisms of νM.

Vector bundles are hence useful (among other things) to understand neighbourhoods of submanifolds. Since we will be interested essentially in manifolds of dimension at most 3, two simple cases will be important.

Proposition 1.1.12. *A connected compact manifold M has a unique line bundle $E \to M$ up to isomorphism with orientable total space E.*

The orientable line bundle on M is a product $M \times \mathbb{R}$ precisely when M is also orientable. If M is not orientable, the unique orientable line bundle is indicated by $M \mathbin{\widetilde{\times}} \mathbb{R}$.

Proposition 1.1.13. *For every n, there are exactly two vector bundles of dimension n over S^1 up to isomorphism, one of which is orientable.*

Again, the orientable vector bundle is $S^1 \times \mathbb{R}^n$ and the non-orientable one is denoted by $S^1 \widetilde{\times} \mathbb{R}^n$. These simple facts allow us to fully understand the possible neighbourhoods of curves in surfaces, and of curves and surfaces inside orientable 3-manifolds.

1.1.11. Manifolds with boundary. Let a *differentiable manifold M with boundary* be a topological space with charts on a fixed closed half-space of $\mathbb{R}^n$ instead of $\mathbb{R}^n$, forming a smooth atlas. (By definition, maps from subsets of $\mathbb{R}^n$ are smooth if they locally admit extensions to smooth functions defined on open domains.)

The points corresponding to the boundary of the half-space form a subset of M denoted by ∂M and called *boundary*. The boundary of a n-manifold is naturally a $(n-1)$-dimensional manifold without boundary. The *interior* of M is $M \setminus \partial M$.

We can define the tangent space $T_p M$ at a point $p \in \partial M$ as the set of all curves in M starting from or arriving to p, with the same equivalence relation as above. The space $T_p M$ is a vector space that contains the hyperplane $T_p \partial M$. Most of the notions introduced for manifolds extend in an appropriate way to manifolds with boundary. A *submanifold* of a manifold with boundary M is the image of an embedding $N \hookrightarrow M$, where N is another manifold with boundary.

The most important manifold with boundary is certainly the *disc*

$$D^n = \left\{ x \mid \|x\| \leqslant 1 \right\} \subset \mathbb{R}^n.$$

More generally, a *disc* in a n-manifold N is a submanifold $D \subset N$ with boundary diffeomorphic to D^n. Since a disc is in fact a (closed) tubular neighbourhood of any point in its interior, the uniqueness of tubular neighbourhoods implies the following.

Theorem 1.1.14 (Cerf – Palais). *Let N be a connected and oriented smooth n-manifold. Two orientation-preserving embeddings $f, g \colon D^n \to \mathrm{int}(N)$ are ambiently isotopic.*

A *boundary component N* of M is a connected component of ∂M. A *collar* for N is an open neighbourhood diffeomorphic to $N \times [0,1)$. As for tubular neighbourhoods, every compact boundary component has a collar, unique up to isotopy.

A *closed manifold* is a compact manifold without boundary.

1.1.12. Cut and paste. If $M \subset N$ is a closed and orientable $(n-1)$-submanifold in the interior of an orientable n-manifold N, it has a tubular neighbourhood diffeomorphic to $M \times \mathbb{R}$. The operation of *cutting N* along M consists of the removal of the portion $M \times (-1, 1)$. The resulting manifold N' has two new boundary components $M \times \{-1\}$ and $M \times \{1\}$, both diffeomorphic to M. By the uniqueness of the tubular neighbourhood, the manifold N' depends (up to diffeomorphisms) only on the isotopy class of $M \subset N$.
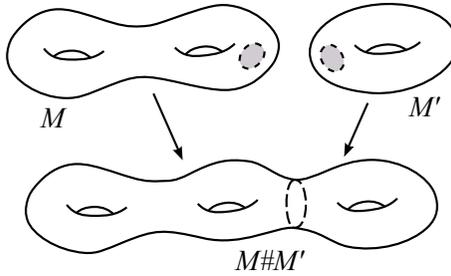
Figure 1.3. A connected sum of closed surfaces.

Let M and N be connected. The submanifold M is *separating* if its complement consists of two connected components. The cut manifold has two or one components depending on whether M is separating or not.

Conversely, let M and N be two n-manifolds with boundary, and let $\varphi\colon \partial M \to \partial N$ be a diffeomorphism. It is possible to *glue* M and N along φ and obtain a new n-manifold as follows.

A naïve approach would consist in taking the topological space $M \sqcup N$ and identify p with $\varphi(p)$ for all $p \in \partial M$. The resulting quotient space is indeed a topological manifold, but the construction of a smooth atlas is not immediate. A quicker method consists of taking two collars $\partial M \times [0, 1)$ and $\partial N \times [0, 1)$ of the boundaries and then consider the topological space

$$(M \setminus \partial M) \sqcup (N \setminus \partial N).$$

Now we identify the points (p, t) and $(\varphi(p), 1 - t)$ of the open collars, for all $p \in \partial M$ and all $t \in (0, 1)$. Having now identified two *open* subsets of $M \setminus \partial M$ and $N \setminus \partial N$, a differentiable atlas for the new manifold is immediately derived from the atlases of M and N.

Proposition 1.1.15. *The resulting smooth manifold depends (up to diffeomorphism) only on the isotopy class of φ.*

1.1.13. Connected sum. We introduce an important cut-and-paste operation. The *connected sum* of two connected oriented n-manifolds M and M' is a new n-manifold obtained by choosing two n-discs $D \subset \mathrm{int}(M)$ and $D' \subset \mathrm{int}(M')$ and an orientation-reversing diffeomorphism $\varphi\colon D \to D'$. The new manifold is constructed in two steps: we first remove the interiors of D and D' from $M \sqcup M'$, thus creating two new boundary components ∂D and $\partial D'$, and then glue these boundary components along $\varphi|_{\partial D}$. See Figure 1.3.

We denote the resulting manifold by $M \# M'$. Since φ is orientation-reversing, the manifold $M \# M'$ is oriented coherently with M and M'. Theorem 1.1.14 implies the following.

Proposition 1.1.16. *The oriented smooth manifold $M\#M'$ depends (up to diffeomorphism) only on M and M'. The connected sum operation $\#$ is commutative, associative, and S^n serves as the identity element.*

We can invert a connected sum as follows. Suppose a n-manifold M contains a separating $(n-1)$-sphere $N \subset \text{int}(M)$. We fix a diffeomorphism $\varphi\colon N \to \partial D^n$. By cutting M along N we get two new boundary components diffeomorphic to N and we glue two discs to them via the map φ. We get two manifolds N_1, N_2 such that $M = N_1\#N_2$.

1.1.14. Transversality. Let $f\colon M \to N$ be a smooth map between manifolds and $X \subset N$ be a submanifold. We say that f is *transverse* to X if for any $p \in f^{-1}X$ the following condition holds:

$$\text{Im}\,(df_p) + T_{f(p)}X = T_{f(p)}N.$$

The maps transverse to a fixed X are generic, that is they form an open dense subset in the space of all smooth maps from X to Y, with respect to an appropriate topology. In particular the following holds.

Theorem 1.1.17. *Let $f\colon M \to N$ be a continuous map and d a distance on N compatible with the topology of N. For every $\varepsilon > 0$ there is a smooth map g transverse to X, homotopic to f, with $d(f(p), g(p)) < \varepsilon$ for all $p \in M$.*

1.2. Riemannian geometry

We briefly introduce Riemannian manifolds and their geometric properties: distance, geodesics, volume, exponential map, injectivity radius, completeness, curvature, and isometries. These notions are carefully explained in do Carmo's *Riemannian Geometry* [**16**].

1.2.1. Metric tensor. A differentiable manifold lacks many natural geometric notions, such as distance between points, angle between vectors, path lengths, geodesics, volumes, etc. It is a quite remarkable fact that the introduction of a single additional mathematical entity suffices to recover all these geometric notions: this miraculous object is the metric tensor.

A *metric tensor* for a differentiable manifold M is the datum of a scalar product on each tangent space T_pM of M, which varies smoothly on p. More specifically, on a chart the scalar product may be expressed as a matrix, and we require that its coefficients vary smoothly on p.

Definition 1.2.1. A *Riemannian manifold* is a differentiable manifold with a metric tensor that is positive definite at every point.

A Riemannian manifold is usually denoted as a pair (M, g), where M is the manifold and g is the metric tensor.

Example 1.2.2. The *Euclidean space* is $\mathbb{R}^n$ with the Euclidean metric tensor $g(x, y) = \sum_{i=1}^{n} x_i y_i$ at every tangent space $T_p\mathbb{R}^n = \mathbb{R}^n$.

Example 1.2.3. Every differential submanifold N in a Riemannian manifold M is also Riemannian: it suffices to restrict at every $p \in N$ the metric tensor on T_pM to the linear subspace T_pN.

In particular, the *sphere*

$$S^n = \left\{ x \in \mathbb{R}^{n+1} \mid \|x\| = 1 \right\}$$

is a submanifold of $\mathbb{R}^{n+1}$ and is hence Riemannian.

The metric tensor g defines in particular a norm for every tangent vector, and an angle between tangent vectors with the same basepoint. The velocity $\gamma'(t)$ of a curve $\gamma \colon I \to M$ at time $t \in I$ now has a norm $\|\gamma'(t)\| \geqslant 0$ called *speed*, and two curves that meet at a point with non-zero velocities form a well-defined angle. The *length* of γ is defined as

$$L(\gamma) = \int_I \|\gamma'(t)\| dt$$

and can be finite or infinite. A *reparametrisation* of γ is the curve $\eta \colon J \to M$ obtained as $\eta = \gamma \circ \varphi$ where $\varphi \colon J \to I$ is a diffeomorphism of intervals. The length is invariant under reparametrisations, that is $L(\gamma) = L(\eta)$.

1.2.2. Distance and geodesics. Let (M, g) be a connected Riemannian manifold. The curves in M now have a length and hence may be used to define a distance on M.

Definition 1.2.4. The *distance* $d(p, q)$ between $p, q \in M$ is

$$d(p, q) = \inf_\gamma L(\gamma)$$

where γ varies among all curves $\gamma \colon [0, 1] \to M$ with $\gamma(0) = p$, $\gamma(1) = q$.

The manifold M equipped with the distance d is a metric space (compatible with the initial topology of M).

Definition 1.2.5. A *geodesic* is a curve $\gamma \colon I \to M$ with constant speed k that realises locally the distance. This means that every $t \in I$ has a closed neighbourhood $[t_0, t_1] \subset I$ with $d(\gamma(t_0), \gamma(t_1)) = L(\gamma|_{[t_0, t_1]}) = k(t_1 - t_0)$.

Note that with this definition the constant curve $\gamma(t) = p_0$ is a geodesic with constant speed $k = 0$. Such a geodesic is called *trivial*. A curve that realises the distance locally may not realise it globally.

Example 1.2.6. The non-trivial geodesics in Euclidean space $\mathbb{R}^n$ are affine lines run at constant speed. The non-trivial geodesics in the sphere S^n are portions of great circles, run at constant speed.

A geodesic $\gamma : I \to M$ is *maximal* if it cannot be extended to a geodesic on a strictly larger interval $J \supset I$. Maximal geodesics are determined by some first-order conditions:

Theorem 1.2.7. *Let $p \in M$ be a point and $v \in T_pM$ a tangent vector. There exists a unique maximal geodesic $\gamma \colon I \to M$ with $\gamma(0) = p$ and $\gamma'(0) = v$. The interval I is open and contains 0.*

1.2.3. Volume. If the differentiable manifold M is oriented, the metric tensor also induces a *volume form*.

The best method to define volumes on a n-manifold M is to construct an appropriate n-form. A n-form ω is an alternating multilinear form

$$\omega_p : \underbrace{T_p M \times \ldots \times T_p M}_{n} \longrightarrow \mathbb{R}$$

at each point $p \in M$, which varies smoothly with p. The alternating condition means that if we swap two vectors the result changes by a sign. The n-forms are designed to be integrated: it makes sense to write

$$\int_D \omega$$

on any open set D. A *volume form* on an oriented manifold M is a n-form ω such that $\omega_p(v_1, \ldots, v_n) > 0$ for every positive basis $v_1, \ldots, v_n$ of T_p and every $p \in M$.

The metric tensor defines a volume form as follows: it suffices to set $\omega_p(e_1, \ldots, e_n) = 1$ on every positive orthonormal basis $e_1, \ldots, e_n$. With this definition every open set D of M has a well-defined *volume*

$$\text{Vol}(D) = \int_D \omega$$

which is a positive number or infinity. If D has compact closure the volume is necessarily finite; in particular, a compact Riemannian manifold M has finite volume $\text{Vol}(M)$.

On a chart, the volume form can be calculated from the metric tensor g via the following formula:

$$\omega = \sqrt{\det g} \cdot dx_1 \cdots dx_n.$$

1.2.4. Completeness. A Riemannian manifold (M, g) is also a metric space, and as such it may be complete or not. For instance, a compact Riemannian manifold is always complete; on the other hand, by removing a point from a Riemannian manifold we always get a non-complete space. Non-compact manifolds like $\mathbb{R}^n$ typically admit both complete and non-complete Riemannian structures. The completeness of a Riemannian manifold may be expressed in various ways:

Theorem 1.2.8 (Hopf – Rinow). *Let (M, g) be a connected Riemannian manifold. The following are equivalent:*

(1) *M is complete,*
(2) *a subset of M is compact if and only if it is closed and bounded,*
(3) *every geodesic can be extended on the whole $\mathbb{R}$.*

If M is complete, any two points $p, q \in M$ are joined by a minimising *geodesic γ, i.e. a curve such that $L(\gamma) = d(p, q)$.*

1.2.5. Exponential map. Let M be a complete Riemannian manifold. The fact that geodesics are determined by first-order conditions allows us to introduce the following.

Definition 1.2.9. Let $p \in M$ be a point. The *exponential map*

$$\exp_p \colon T_p M \longrightarrow M$$

is defined as follows. A vector $v \in T_p M$ determines a maximal geodesic $\gamma_v \colon \mathbb{R} \to M$ with $\gamma_v(0) = p$ and $\gamma'_v(0) = v$. We set $\exp_p(v) = \gamma_v(1)$.

Theorem 1.2.10. *The differential of the exponential map $\exp_p$ at the origin is the identity and hence $\exp_p$ is a local diffeomorphism at the origin.*

Via the exponential map, a small neighborhood of the origin in $T_p M$ can be used as a chart near p: we recover here the intuitive idea that the tangent space approximates the manifold near p.

When M is not complete, the exponential map is only defined in some open star-shaped neighbourhood $V_p \subset T_p M$ of the origin, and Theorem 1.2.10 holds also in this case.

1.2.6. Injectivity radius. Let M be a complete Riemannian manifold. We introduce the following.

Definition 1.2.11. The *injectivity radius* $\mathrm{inj}_p M > 0$ of M at a point $p \in M$ is the supremum of all $r > 0$ such that $\exp_p |_{B_0(r)}$ is a diffeomorphism onto its image.

Here $B_0(r)$ is the open ball with centre 0 and radius r in the tangent space $T_p M$. The injectivity radius is positive by Theorem 1.2.10. For every $r < \mathrm{inj}_p M$ the exponential map transforms the ball of radius r in $T_p(M)$ into the metric ball of radius r in M. That is, the following equality holds:

$$\exp_p(B_0(r)) = B_p(r)$$

and the metric ball $B_p(r)$ is indeed diffeomorphic to an open ball in $\mathbb{R}^n$. When $r \geq \mathrm{inj}_p M$ this may not be true: for instance if M is compact there is a $R > 0$ such that $B_p(R) = M$.

The injectivity radius $\mathrm{inj}_p(M)$ varies continuously with respect to $p \in M$; the injectivity radius $\mathrm{inj}(M)$ of M is defined as

$$\mathrm{inj}(M) = \inf_{p \in M} \mathrm{inj}_p M.$$

Proposition 1.2.12. *A compact Riemannian manifold has positive injectivity radius.*

Proof. The injectivity radius $\mathrm{inj}_p M$ is positive and varies continuously with p, hence it has a positive minimum on the compact M.   □

Finally we note the following. A *loop* is a curve $\gamma \colon [a, b] \to M$ with $\gamma(a) = \gamma(b)$.

Proposition 1.2.13. *Let M be a complete Riemannian manifold. Every loop γ in M shorter than $2 \cdot \mathrm{inj}(M)$ is homotopically trivial.*

Proof. Set $p = \gamma(a) = \gamma(b)$. Since γ is shorter than $2 \cdot \mathrm{inj}(M)$, it cannot escape the ball $B_p(r)$ for some $r < \mathrm{inj}(M) \leqslant \mathrm{inj}_p M$. This ball is diffeomorphic to a ball in $\mathbb{R}^n$, hence in particular it is contractible, so γ is homotopically trivial. $\qquad\square$

A proof of the continuity of $\mathrm{inj}_p(M)$ can be found in [**35**].

1.2.7. Curvature. The *curvature* of a Riemannian manifold (M, g) is some mathematical entity that measures how distorted g is when compared to the familiar Euclidean structure on $\mathbb{R}^n$. The curvature is encoded by various kinds of mathematical objects, and some of them may be frustratingly complicated when $n = \dim M \geqslant 3$. Luckily, most of the Riemannian manifolds that we will encounter in this book have "constant curvature" and these objects will simplify dramatically. We briefly recall them here.

The metric tensor g induces the *Levi-Civita connection* ∇, which can be used to differentiate any vector field with respect to a tangent vector at any point of M. We concentrate on an open chart and suppose that $M = U$ is an open set of $\mathbb{R}^n$; let $e_1, \ldots, e_n$ be the coordinate vector fields on U and ∇_i be the Levi-Civita differentiation with respect to e_i. We have

$$\nabla_i e_j = \Gamma_{ij}^k e_k$$

for some smooth functions $\Gamma_{ij}^k \colon U \to \mathbb{R}$ called the *Chistoffel symbols*. Here we use the Einstein notation: repeated indices should be added from 1 to n. The Christoffel symbols are determined by g via the formula:

$$\Gamma_{ij}^k = \frac{1}{2} g^{k\ell} \left(\frac{\partial g_{\ell i}}{\partial x^j} + \frac{\partial g_{\ell j}}{\partial x^i} - \frac{\partial g_{ij}}{\partial x^\ell} \right).$$

Here $g^{k\ell}$ is just the inverse matrix of $g = g_{k\ell}$. Christoffel symbols are very useful but depend heavily on the chosen chart; a more intrinsic object is the *Riemann tensor* that may be defined (quite obscurely) as

$$R_{jk\ell}^i = \partial_k \Gamma_{\ell j}^i - \partial_\ell \Gamma_{kj}^i + \Gamma_{km}^i \Gamma_{\ell j}^m - \Gamma_{\ell m}^i \Gamma_{kj}^m.$$

By contracting two indices we get the *Ricci tensor*

$$R_{ij} = R_{ikj}^k = \partial_k \Gamma_{ji}^k - \partial_j \Gamma_{ki}^k + \Gamma_{km}^k \Gamma_{ji}^m - \Gamma_{jm}^k \Gamma_{ki}^m.$$

The Ricci tensor is symmetric, and by further contracting we get the *scalar curvature*

$$R = g^{ij} R_{ij}.$$
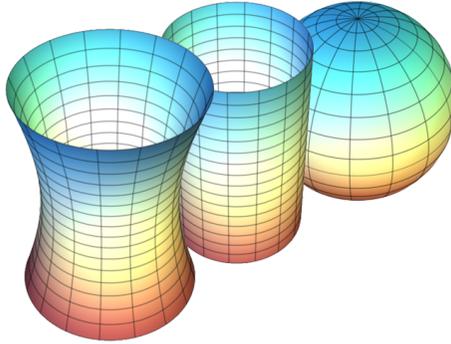
Figure 1.4. Three surfaces in space (hyperboloid with one sheet, cylinder, sphere) whose gaussian curvature is respectively negative, null, and positive at each point. The curvature on the sphere is constant, while the curvature on the hyperboloid varies: a complete surface in $\mathbb{R}^3$ cannot have constant negative curvature.

1.2.8. Scalar curvature. What kind of geometric information is contained in the tensors introduced above? The scalar curvature R is the simplest curvature detector, and it measures the local discrepancy between volumes in M and in $\mathbb{R}^n$. Let $B_p(\varepsilon) \subset M$ be the ε-ball in M centred at p and $B_0(\varepsilon) \subset \mathbb{R}^n$ be the ε-ball in the Euclidean $\mathbb{R}^n$. We have

$$\mathrm{Vol}\big(B_p(\varepsilon)\big) = \mathrm{Vol}\big(B_0(\varepsilon)\big) \cdot \left(1 - \frac{R}{6(n+2)}\varepsilon^2\right) + \mathrm{O}(\varepsilon^4).$$

We note in particular that R is positive (negative) if $B_p(\varepsilon)$ has smaller (bigger) volume than the usual Euclidean volume.

If M has dimension 2, that is it is a surface, the Riemann and Ricci tensors simplify dramatically and are fully determined by the scalar curvature R, which is in turn equal to twice the *gaussian curvature* K: if M is contained in $\mathbb{R}^3$ the gaussian curvature is defined as the product of its two principal curvatures, but when M is abstract principal curvatures make no sense. On surfaces, the formula above reads as

$$\mathrm{Area}(B_p(\varepsilon)) = \pi\varepsilon^2 - \frac{\pi\varepsilon^4}{12}K + \mathrm{O}(\varepsilon^4).$$

1.2.9. Sectional curvature. If (M, g) has dimension $n \geqslant 3$ the scalar curvature is a weak curvature detector when compared to the Ricci and Riemann tensors. Moreover, there is yet another curvature detector which encodes the same amount of information of the full Riemann tensor, but in a more geometric way: this is the *sectional curvature*.

Definition 1.2.14. Let (M, g) be a Riemannian manifold. Let $p \in M$ be a point and $W \subset T_pM$ be a 2-dimensional vector subspace. By Theorem

1.2.10 there exists an open neighbourhood $U_p \subset T_p M$ of the origin where $\exp_p$ is a diffeomorphism onto its image. In particular $S = \exp_p(U_p \cap W)$ is a smooth surface in M passing through p, with tangent plane W. As a submanifold of M, the surface S inherits a Riemannian structure from g.

The *sectional curvature* of (M, g) along W is defined as the gaussian curvature of S in p.

We can use the sectional curvature to unveil the geometric nature of the Ricci tensor: the Ricci tensor R_{ij} in p measures the average sectional curvature along axes, that is for every unit vector $v \in T_p(M)$ the number $R_{ij} v^i v^j$ is $(n-1)$ times the average sectional curvature of the planes $W \subset T_p(M)$ containing v.

The Riemann tensor is determined by the sectional curvatures and vice-versa. In dimension $n = 3$ the Ricci tensor fully determines the sectional curvatures and hence also the Riemann tensor. This is not true in dimension $n \geqslant 4$.

1.2.10. Constant sectional curvature. A Riemannian manifold (M, g) has *constant sectional curvature K* if the sectional curvature of every 2-dimensional vector space $W \subset T_p M$ at every point $p \in M$ is always K.

Remark 1.2.15. On a Riemannian manifold (M, g) one may *rescale* the metric by some factor $\lambda > 0$ substituting g with the tensor λg. At every point the scalar product is rescaled by λ. Consequently, lengths of curves are rescaled by $\sqrt{\lambda}$ and volumes are rescaled by $\lambda^{\frac{n}{2}}$. The sectional curvature is rescaled by $1/\lambda$.

By rescaling the metric we may transform every Riemannian manifold with constant sectional curvature K into one with constant sectional curvature -1, 0, or 1.

Example 1.2.16. Euclidean space $\mathbb{R}^n$ has constant curvature zero. A sphere of radius R has constant curvature $1/R^2$.

1.2.11. Isometries. Every honest category has its morphisms. Riemannian manifolds are so rigid, that in fact one typically introduces only isomorphisms: these are called *isometries*.

Definition 1.2.17. A diffeomorphism $f : M \to N$ between two Riemannian manifolds (M, g) and (N, h) is an *isometry* if it preserves the scalar product. That is, the equality

$$\langle v, w \rangle = \langle df_p(v), df_p(w) \rangle$$

holds for all $p \in M$ and every pair of vectors $v, w \in T_p M$. The symbols $\langle , \rangle$ indicate the scalar products in $T_p M$ and $T_{f(p)} N$.

As we said, isometries are extremely rigid. These are determined by their first-order behaviour at any single point.

Theorem 1.2.18. *Let $f, g\colon M \to N$ be two isometries between two connected Riemannian manifolds. If there is a point $p \in M$ such that $f(p) = g(p)$ and $df_p = dg_p$, then $f = g$ everywhere.*

Proof. Let us show that the subset $S \subset M$ of the points p such that $f(p) = g(p)$ and $df_p = dg_p$ is open and closed.

The locus where two functions coincide is typically closed, and this holds also here (to prove it, take a chart). We prove that it is open: pick $p \in S$. By Theorem 1.2.10 there is an open neighbourhood $U_p \subset T_p M$ of the origin where the exponential map is a diffeomorphism onto its image. We show that the open set $\exp_p(U_p)$ is entirely contained in S.

A point $q \in \exp_p(U_p)$ is the image $q = \exp(v)$ of a vector $v \in U_p$ and hence $q = \gamma(1)$ for the geodesic γ determined by the data $\gamma(0) = p, \gamma'(0) = v$. The maps f and g are isometries and hence send geodesics to geodesics: here $f \circ \gamma$ and $g \circ \gamma$ are geodesics starting from $f(p) = g(p)$ with the same initial velocities and thus they coincide. This implies that $f(q) = g(q)$. Since f and g coincide on the open set $\exp_p(U_p)$, also their differentials do. $\square$

1.2.12. Local isometries. A *local isometry* $f\colon M \to N$ between Riemannian manifolds is a map where every $p \in M$ has an open neighbourhood U such that $f|_U$ is an isometry onto its image. Theorem 1.2.18 applies with the same proof to local isometries.

The following proposition relates nicely the notions of local isometry, topological covering, and completeness.

Proposition 1.2.19. *Let $f\colon M \to N$ be a local isometry.*
  (1) *If M is complete, the map f is a covering.*
  (2) *If f is a covering, then M is complete $\iff N$ is complete.*

Proof. Since f is a local isometry, every geodesic in M projects to a geodesic in N. If f is also a covering, the converse holds: every geodesic in N lifts to a geodesic in M (at any starting point).

If f is a covering we can thus project and lift geodesics via f: therefore every geodesic in M can be extended to $\mathbb{R}$ if and only if every geodesic in N can; this proves (2) using the Hopf $-$ Rinow Theorem 1.2.8.

We prove (1) by showing that the ball $B = B(p, \mathrm{inj}_p N)$ is well-covered for all $p \in N$. Since M is complete, every geodesic in N can be lifted to a geodesic in M (at any starting point). For every $\tilde{p} \in f^{-1}(p)$ the map f sends the geodesics exiting from $\tilde{p}$ to geodesics exiting from p and hence sends isometrically $B(\tilde{p}, \mathrm{inj}_p N)$ onto B. On the other hand, given a point $q \in f^{-1}(B)$, the geodesic in B connecting $f(q)$ to p lifts to a geodesic connecting q to some point $\tilde{p} \in f^{-1}(p)$. Therefore

$$f^{-1}\big(B(p, \mathrm{inj}_p N)\big) = \bigsqcup_{\tilde{p} \in f^{-1}(p)} B(\tilde{p}, \mathrm{inj}_p N)$$

and f is a covering. $\square$

Proposition 1.2.20. *Let* $f : M \to N$ *be a local isometry and a degree-d covering. We have*

$$\mathrm{Vol}(M) = d \cdot \mathrm{Vol}(N).$$

Sketch of the proof. We may find a disjoint union of well-covered open sets in N whose complement has zero measure. Every such open set lifts to d copies of it in M, and the zero-measure set lifts to a zero-measure set. $\qquad\square$

The formula makes sense also when some of the quantities $\mathrm{Vol}(M)$, $\mathrm{Vol}(N)$, and d are infinite.

1.2.13. Totally geodesic submanifolds. A differentiable submanifold M in a Riemannian manifold N is *totally geodesic* if every geodesic in M with the induced metric is also a geodesic in N.

When dim $M = 1$ this notion is equivalent to that of an unparametrized embedded geodesic; if dim $M \geqslant 2$ then M is a quite peculiar object: generic Riemannian manifolds do not contain totally geodesic surfaces at all. An equivalent condition is that, for every $p \in M$ and every $v \in T_p(M)$, the unique geodesic in N passing through p with velocity v stays in M for some interval $(-\varepsilon, \varepsilon)$.

1.2.14. Riemannian manifolds with boundary. Many geometric notions in Riemannian geometry extend easily to manifolds M with boundary. The boundary ∂M of a Riemannian manifold M is naturally a Riemannian manifold without boundary. A particularly nice (and exceptional) case is when ∂M is totally geodesic.

1.3. Measure theory

We will use some basic measure theory only in Chapter 8.

1.3.1. Borel measure. A *Borel set* in a topological space X is any subset obtained from open sets through the operations of countable union, countable intersection, and relative complement. Let $\mathscr{F}$ denote the set of all Borel sets. A *Borel measure* on X is a function $\mu \colon \mathscr{F} \to [0, +\infty]$ which is additive on any countable collection of disjoint sets.

The measure is *locally finite* if every point has a neighbourhood of finite measure and is *trivial* if $\mu(S) = 0$ for all $S \in \mathscr{F}$.

Exercise 1.3.1. If μ is a locally finite Borel measure then $\mu(X) < +\infty$ for any compact Borel set $K \subset X$.

Example 1.3.2. Let $D \subset X$ be a discrete set. The *Dirac measure* δ_D concentrated in D is the measure

$$\delta_D(S) = \#(S \cap D).$$

Since D is discrete, the measure δ_D is locally finite.

The *support* of a measure is the set of all points $x \in X$ such that $\mu(U) > 0$ for any open set U containing x. The support is a closed subset of X. The measure is *fully supported* if its support is X. The support of δ_D is of course D. A point $x \in X$ is an *atom* for μ if $\mu(\{x\}) > 0$.

1.3.2. Construction by local data. A measure can be defined using local data in the following way.

Proposition 1.3.3. *Let $\{U_i\}_{i \in I}$ be a countable, locally finite open covering of X and for any $i \in I$ let μ_i be a locally finite Borel measure on U_i. If $\mu_i|_{U_i \cap U_j} = \mu_j|_{U_i \cap U_j}$ for all $i, j \in I$ there is a unique locally finite Borel measure μ on X whose restriction to U_i is μ_i for all i.*

Proof. For every finite subset $J \subset I$ we define

$$X_J = \left(\cap_{j \in J} U_j \right) \setminus \left(\cup_{i \in I \setminus J} U_i \right).$$

The sets X_J form a countable partition of X into Borel sets and every X_J is equipped with a measure $\mu_J = \mu_j|_{X_J}$ for any $j \in J$. Define μ by setting

$$\mu(S) = \sum_{j \in J} \mu(S \cap X_j)$$

on any Borel $S \subset X$. $\square$

When X is a reasonable space some hypothesis may be dropped.

Proposition 1.3.4. *If X is paracompact and separable, Proposition 1.3.3 holds for any open covering $\{U_i\}_{i \in I}$.*

Proof. By paracompactness and separability the open covering $\{U_i\}$ has a refinement that is locally finite and countable: apply Proposition 1.3.3 to the refinement to get a measure μ. To prove that indeed $\mu|_{U_i} = \mu_i$ apply Proposition 1.3.3 again to the covering of U_i given by the refinement. $\square$

1.3.3. Topology on the measures space. In what follows we suppose for simplicity that X is a finite-dimensional topological manifold, although everything is valid in a much wider generality. We indicate by $\mathscr{M}(X)$ the space of all locally finite Borel measures on X and by $C_c(X)$ the space of all continuous functions $X \to \mathbb{R}$ with compact support: the space $C_c(X)$ is not a Banach space, but it is a topological vector space.

Recall that the *topological dual* of a topological vector space V is the vector space V^* formed by all continuous linear functionals $V \to \mathbb{R}$. A measure $\mu \in \mathscr{M}(X)$ acts like a continuous functional on $C_c(X)$ as follows

$$\mu : f \longmapsto \int_\mu f$$

and hence defines an element of $C_c^*(X)$. A functional in $C_c^*(X)$ is *positive* if it assumes non-negative values on non-negative functions.

Theorem 1.3.5 (Riesz representation). *The space $\mathscr{M}(X)$ may be identified in this way to the subset in $C_c(X)^*$ of all positive functionals.*

The space $\mathscr{M}(X)$ in $C_c(X)^*$ is closed with respect to sum and product with a positive scalar. We now use the embedding of $\mathscr{M}(X)$ into $C_c(X)^*$ to define a natural topology on $\mathscr{M}(X)$.

Definition 1.3.6. Let V be a real topological vector space. Every vector $v \in V$ defines a functional in V^* as $f \mapsto f(v)$. The *weak-* topology* on V^* is the weakest topology among those where these functionals are continuous.

We give $C_c(X)^*$, and hence $\mathscr{M}(X)$, the weak-* topology.

1.3.4. Sequences of measures. By definition, a sequence of measures μ_i converges to μ if and only if $\int_{\mu_i} f \to \int_\mu f$ for any $f \in C_c(X)$. This type of weak convergence is usually denoted with the symbol $\mu_i \rightharpoonup \mu$.

Exercise 1.3.7. Let x_i be a sequence of points in X that tend to $x \in X$: we get $\delta_{x_i} \rightharpoonup \delta_x$.

It is important to note that $\mu_i \rightharpoonup \mu$ does *not* imply $\mu_i(U) \to \mu(U)$ for any open (or closed) set U: consider for instance a sequence $\mu_i = \delta_{x_i}$ with x_i exiting from (or entering into) the set U. We can get this convergence on compact sets if we can control their topological boundary.

Proposition 1.3.8. *Let $K \subset X$ be a Borel compact subset. If $\mu_i \rightharpoonup \mu$ and $\mu(\partial K) = 0$ then $\mu_i(K) \to \mu(K)$.*

On Banach spaces, the unitary ball is compact in the weak-* topology. Here $\mathscr{M}(X)$ is not a Banach space, but we have an analogous compactness theorem.

Theorem 1.3.9. *A sequence of measures μ_i such that $\mu_i(K)$ is bounded on every Borel compact set $K \subset X$ converges on a subsequence.*

1.4. Groups

We recall some basic definitions and properties of groups.

1.4.1. Presentations. Recall that a *finite presentation* of a group G is a description of G as

$$\langle g_1, \ldots, g_k \mid r_1, \ldots, r_s \rangle$$

where $g_1, \ldots, g_k \in G$ are the *generators* and $r_1, \ldots, r_s$ are words in $g_i^{\pm 1}$ called *relations*, such that

$$G \cong F(g_i)/_{N(r_j)}$$

where $F(g_i)$ is the free group generated by the g_i's and $N(r_j) \triangleleft F(g_i)$ is the *normal closure* of the r_j's, the smallest normal subgroup containing them.

Not every group G has a finite presentation: a necessary (but not sufficient) condition is that G must be finitely generated.

1.4.2. Commutators. Let G be a group. The *commutator* of two elements $h, k \in G$ is the element

$$[h, k] = hkh^{-1}k^{-1}.$$

The commutator $[h, k]$ is trivial if and only if h and k commute. Commutators do not form a subgroup in general, but we can use them to generate one.

More generally, let $H, K < G$ be two subgroups. We define $[H, K]$ as the subgroup of G generated by all commutators $[h, k]$ where h and k vary in H and K, respectively. Every element in $[H, K]$ is a product of commutators $[h, k]$ and of their inverses $[h, k]^{-1} = [k, h]$, and we have $[H, K] = [K, H]$.

Proposition 1.4.1. *If H and K are normal subgroups of G then $[H, K] < H \cap K$ and $[H, K]$ is a normal subgroup of G.*

Proof. If H and K are normal we get $[h, k] = hkh^{-1}k^{-1} \in H \cap K$. Moreover, for every $g \in G$ we have

$$g^{-1} \cdot [h_1, k_1]^{\pm 1} \cdots [h_i, k_i]^{\pm 1} \cdot g = [g^{-1}h_1 g, g^{-1}k_1 g]^{\pm 1} \cdots [g^{-1}h_i g, g^{-1}k_i g]^{\pm 1}$$

hence $[H, K]$ is normal in G if H and K are.   □

The group $[G, G]$ is the *commutator subgroup* of G. It is trivial if and only if G is abelian.

1.4.3. Series. We can use commutators iteratively to create some characteristic subgroups of G. There are two natural ways to do this, and they produce two nested sequences of subgroups.

The *lower central series* of G is the sequence of normal subgroups

$$G = G_0 > G_1 > \ldots > G_n > \ldots$$

defined iteratively by setting

$$G_{n+1} = [G_n, G].$$

The *derived series* is the sequence of normal subgroups

$$G = G^{(0)} \triangleright G^{(1)} \triangleright \ldots \triangleright G^{(n)} \triangleright \ldots$$

defined by setting

$$G^{(n+1)} = [G^{(n)}, G^{(n)}].$$

We clearly have $G^{(n)} < G_n$ for all n.

Definition 1.4.2. A group G is *nilpotent* if G_n is trivial for some n. It is *solvable* if $G^{(n)}$ is trivial for some n.

The following implications are obvious:

$$G \text{ abelian} \implies G \text{ nilpotent} \implies G \text{ solvable}.$$

The converses are false, as we will see.

Proposition 1.4.3. *Subgroups and quotients of abelian (nilpotent, solvable) groups are also abelian (nilpotent, solvable).*

Proof. If $H < G$ then $H_n < G_n$ and $H^{(n)} < G^{(n)}$. If $H = G/_N$, then $H_n = G_n/_N$ and $H^{(n)} = G^{(n)}/_N$. This is proved by induction on n. $\square$

1.4.4. Nilpotent groups. Every abelian group G is obviously nilpotent since G_1 is trivial.

Exercise 1.4.4. Let the *Heisenberg group* Nil consist of all matrices

$$\begin{pmatrix} 1 & x & z \\ 0 & 1 & y \\ 0 & 0 & 1 \end{pmatrix}$$

where x, y, and z vary in $\mathbb{R}$, with the multiplication operation. Prove that the Heisenberg group is non-abelian and nilpotent. Indeed the matrices with $x = y = 0$ form an abelian subgroup $\mathbb{R} < $ Nil and we have

$$[\text{Nil}, \text{Nil}] = \mathbb{R}, \qquad [\mathbb{R}, \text{Nil}] = \{e\}.$$

We will use at some point the following criterion.

Proposition 1.4.5. *Let G be a group generated by some set S and $n > 0$ a number. Suppose that*

$$[a_1, \ldots [a_{n-1}, [a_n, b]] \cdots]$$

is trivial for all $a_1, \ldots, a_n, b \in S$. Then $G_n = \{e\}$ and thus G is nilpotent.

Proof. We claim that G_n is generated by some elements of type

$$[a_1, \ldots [a_{m-1}, [a_m, b]] \cdots]$$

with $m \geqslant n$ and $a_1, \ldots, a_m, b \in S$: this clearly implies the proposition. The claim is proved by induction on n using the formula

$$[a, bc] = [a, b] \cdot [b, [a, c]] \cdot [a, c]$$

which holds in every group. $\square$

We note the following.

Proposition 1.4.6. *A nilpotent non-trivial group has non-trivial centre.*

Proof. Let G_n be the last non-trivial group in the lower central series. Since $G_{n+1} = [G_n, G]$ is trivial, the centre of G contains G_n. $\square$

1.4.5. Solvable groups. Solvable groups form a strictly larger class than nilpotent groups.

Exercise 1.4.7. The permutation group S_3 is solvable but not nilpotent.

Exercise 1.4.8. The group $\mathrm{Aff}(\mathbb{R}) = \{x \mapsto ax + b \mid a \in \mathbb{R}^*, b \in \mathbb{R}\}$ of affine transformations is solvable but not nilpotent. Indeed the subgroup $[\mathrm{Aff}(\mathbb{R}), \mathrm{Aff}(\mathbb{R})] = \mathbb{R}$ consists of all translations and

$$[\mathbb{R}, \mathrm{Aff}(\mathbb{R})] = \mathbb{R}, \qquad [\mathbb{R}, \mathbb{R}] = \{e\}.$$

Exercise 1.4.9. The permutation group S_n is not solvable for $n \geqslant 5$.

The latter fact is related to the existence of polynomials of degree $n \geqslant 5$ that are not solvable by radicals.

Solvable groups are farther from being abelian than nilpotent groups: for instance, they may have trivial centre (like S_3). However, they still share some nice properties with the abelian world:

Proposition 1.4.10. *A solvable group contains a non-trivial normal abelian subgroup.*

Proof. Let $G^{(n)}$ be the last non-trivial group in the derived series. It is normal in G and abelian, since $[G^{(n)}, G^{(n)}]$ is trivial.   □

1.4.6. Lie groups. A *Lie group* is a smooth manifold G which is also a group, such that the operations

$$G \times G \to G, \quad (a, b) \mapsto ab$$
$$G \to G, \quad a \mapsto a^{-1}$$

are smooth.

Example 1.4.11. The basic examples are $\mathrm{GL}(n, \mathbb{R})$ and $\mathrm{GL}(n, \mathbb{C})$ consisting of invertible real and complex $n \times n$ matrices, with the multiplication operation. These contain many interesting Lie subgroups:

- $\mathrm{SL}(n, \mathbb{R})$ and $\mathrm{SL}(n, \mathbb{C})$, the matrices with determinant 1;
- $\mathrm{O}(n)$ and $\mathrm{U}(n)$, the orthogonal and unitary matrices;
- $\mathrm{SO}(n) = \mathrm{SL}(n, \mathbb{R}) \cap \mathrm{O}(n)$ and $\mathrm{SU}(n) = \mathrm{SL}(n, \mathbb{C}) \cap \mathrm{U}(n)$.

The Lie groups $\mathrm{O}(n)$ and $\mathrm{U}(n)$ are compact.

Exercise 1.4.12. The following Lie groups are isomorphic:

$$S^1 \cong \mathrm{U}(1) \cong \mathrm{SO}(2).$$

Here are some ways to construct more examples:

- the product of two Lie groups is a Lie group,
- a closed subgroup of a Lie group is a Lie group,
- the universal cover of a Lie group is a Lie group.

A Lie group G is often not connected, and we denote by G° the connected component of G containing the identity. The subset G° is a normal Lie subgroup of G. For instance $O(n)$ has two components and $O(n)^\circ = SO(n)$.

Another important example is the Lie group

$$O(m, n) = \left\{ A \in GL(m + n, \mathbb{R}) \mid {}^t\!A I_{m,n} A = I_{m,n} \right\}$$

defined for any pair $m, n \geqslant 0$ of integers, where $I_{m,n}$ is the diagonal matrix

$$I_{m,n} = \begin{pmatrix} I_m & 0 \\ 0 & -I_n \end{pmatrix}.$$

The Lie group $O(m, n)$ has two connected components if either m or n is zero (in that case we recover the orthogonal group) and has four connected components otherwise.

Remark 1.4.13. Not all the Lie groups are subgroups of $GL(n, \mathbb{C})$: for instance, the universal cover of $SL(2, \mathbb{R})$ is not.

A *homomorphism* $\varphi\colon G \to H$ of Lie groups is a smooth group homomorphism. The kernel of φ is a closed subgroup and hence a Lie subgroup of G. An isomorphism of Lie groups is a group isomorphism that is also a diffeomorphism.

1.4.7. Vector fields, metrics, and differential forms. A Lie group G acts on itself in two ways: by left and right multiplication. Both actions are smooth, free, and transitive.

A geometric object on G is *left-invariant* if it is invariant by left multiplication. Right-invariance is defined analogously. In general, it is easy to find objects that are either left- or right-invariant, but not necessarily both.

There is a natural 1-1 correspondence between:

- vectors in $T_e G$ and left-invariant vector fields on G,
- scalar products on $T_e G$ and left-invariant Riemannian metrics on G,
- n-forms on $T_e G$ and left-invariant n-forms on G,
- orientations on $T_e G$ and left-invariant orientations on G.

This holds because left-multiplication by an element $g \in G$ is a diffeomorphism that transports everything from $T_e G$ to $T_g G$. The same holds for right-invariant objects.

Every basis of $T_e G$ extends in this way to n independent left-invariant vector fields: this shows that G is *parallelizable*, *i.e.* the tangent bundle of G is trivial.

1.4.8. Simple Lie groups. A *simple Lie group* G is a connected, non-abelian Lie group G that does not contain any non-trivial connected normal subgroup (the trivial cases being of course $\{e\}$ and G itself).

The classification of simple Lie groups is due to E. Cartan. The following theorem furnishes many examples.

Theorem 1.4.14. *The groups* $O(m, n)^{\circ}$ *are all simple when* $m + n \geqslant 3$, *except the cases* $(m, n) = (4, 0), (2, 2), (0, 4)$.

1.4.9. Haar measures. Let G be a Lie group. As we just said, non-vanishing left-invariant n-forms ω on G are in 1-1 correspondence with non-trivial n-forms on $T_e G$, and since the latter are all proportional, the form ω is unique up to multiplication by a non-zero scalar.

The form ω defines a left-invariant orientation (a basis $v_1, \ldots, v_n$ is positive when $\omega(v_1, \ldots, v_n) > 0$); it is a volume form and hence defines a left-invariant locally finite Borel measure on G, called the *Haar measure* of G. Summing up, the Haar measure depends only on G up to rescaling.

Since left- and right-multiplications commute, right-multiplication by an element $g \in G$ transforms ω into another non-vanishing left-invariant form which must be equal to $\lambda(g)\omega$ for some positive real number $\lambda(g)$. This defines a Lie group homomorphism $\lambda \colon G \to \mathbb{R}_{>0}$ to the multiplicative group of positive real numbers, called the *modular function*.

The Lie group G is *unimodular* if the left-invariant Haar measures are also right-invariant, that is if the modular function λ is trivial. Recall that a group G is *simple* if its normal subgroups are $\{e\}$ and G.

Proposition 1.4.15. *Compact, abelian, and simple Lie groups are unimodular.*

Proof. If G is compact, the image of λ is compact in $\mathbb{R}_{>0}$ and hence trivial. If G is abelian, left- and right-multiplications coincide. If G is simple, the connected normal Lie subgroup $(\ker \lambda)^{\circ}$ must be G. $\square$

Corollary 1.4.16. *The group* $O(m, n)$ *is unimodular if* (m, n) *is distinct from* $(4, 0), (2, 2), (0, 4)$.

Proof. The identity component $O(m, n)^{\circ}$ is simple and hence unimodular. The modular function $\lambda \colon O(m, n) \to \mathbb{R}$ is trivial on the finite-index subgroup $O(m, n)^{\circ}$ and is hence trivial. $\square$

Remark 1.4.17. When $(m, n) = (4, 0), (2, 2), (0, 4)$ the group $O(m, n)$ is *semisimple*, which means that it looks roughly like a product of simple groups, and is in fact unimodular also in this case. We will not need to introduce this concept rigorously here.

The solvable Lie group $\mathrm{Aff}(\mathbb{R})$ is not unimodular.

1.4.10. Discrete subgroups. Let G be a Lie group. A closed subgroup $H < G$ is *discrete* if it forms a discrete topological subset, that is if every point in H is isolated.

Exercise 1.4.18. A subgroup $H < G$ is discrete if and only if $e \in G$ is an isolated point in G.

1.4.11. The Selberg Lemma. The group $GL(n, \mathbb{C})$ is the prototypical Lie group, since it contains many Lie groups. A couple of purely algebraic facts about this group will have some important geometric consequences in this book. The first is the Selberg Lemma. Recall that a group *has no torsion* if every non-trivial element in it has infinite order.

Lemma 1.4.19 (Selberg's Lemma). *Every finitely generated subgroup $G < GL(n, \mathbb{C})$ has a finite-index normal subgroup $H \triangleleft G$ without torsion.*

The second fact is quite related. A group G is *residually finite* if one of the following equivalent conditions holds:

- for every non-trivial element $g \in G$ there is a finite group F and a surjective homomorphism $\varphi \colon G \to F$ with $\varphi(g) \neq e$;
- for every non-trivial element $g \in G$ there is a finite-index normal subgroup $H \triangleleft G$ which does not contain g;
- the intersection of all finite-index normal subgroups in G is trivial.

Lemma 1.4.20. *Every finitely generated subgroup of $GL(n, \mathbb{C})$ is residually finite.*

The proof of these two lemmas is not particularly hard (see for instance [**49**, Chapter 7.6]) but it employs some purely algebraic techniques that are distant from the scope of this book. Their geometric consequences, as we will see, are quite remarkable.

1.4.12. Lie algebras. A *Lie algebra* $\mathfrak{g}$ is a real vector space equipped with an alternate bilinear product

$$[,] \colon \mathfrak{g} \times \mathfrak{g} \to \mathfrak{g}$$

called *Lie bracket* that satisfies the Jacobi identity

$$[x, [y, z]] + [y, [z, x]] + [z, [x, y]] = 0$$

for all $x, y, z \in \mathfrak{g}$.

The tangent space $T_e G$ at $e \in G$ of a Lie group G has a natural Lie algebra structure and is denoted by $\mathfrak{g}$. It is equipped with a natural *exponential map* $\exp \colon \mathfrak{g} \to G$ that sends 0 to e and is a local diffeomorphism at 0.

The differential $d\varphi_e \colon \mathfrak{g} \to \mathfrak{h}$ of a Lie group homomorphism $\varphi \colon G \to H$ is a Lie algebra homomorphism and is denoted by φ_*. The diagram

$$
\begin{array}{ccc}
\mathfrak{g} & \xrightarrow{\varphi_*} & \mathfrak{h} \\
{\scriptstyle \exp} \downarrow & & \downarrow {\scriptstyle \exp} \\
G & \xrightarrow{\varphi} & H
\end{array}
$$

commutes. The basic example is the following: the Lie algebra $\mathfrak{gl}(n, \mathbb{C})$ of $GL(n, \mathbb{C})$ is the vector space $M(n, \mathbb{C})$ of all $n \times n$ matrices; here the map $\exp$ is the usual matrix exponential and $[A, B] = AB - BA$ is the

usual commutator bracket. The Lie algebras of $GL(n, \mathbb{R})$, $SL(n, \mathbb{R})$, $O(n)$, $SO(n)$ are:

$$\mathfrak{gl}(n, \mathbb{R}) = M(n, \mathbb{R}),$$
$$\mathfrak{sl}(n, \mathbb{R}) = \{A \in M(n, \mathbb{R}) \mid \mathrm{tr}A = 0\},$$
$$\mathfrak{o}(n) = \{A \in M(n, \mathbb{R}) \mid A + {}^{t}A = 0\},$$
$$\mathfrak{so}(n) = \{A \in M(n, \mathbb{R}) \mid A + {}^{t}A = 0, \ \mathrm{tr}A = 0\}.$$

Proposition 1.4.21. *Let G and H be connected. A Lie group homomorphism $\varphi \colon G \to H$ is a topological covering if and only if φ_{*} is an isomorphism.*

1.5. Group actions

Groups acting of spaces are so important in geometry, that they are sometimes used as a definition of "geometry" itself.

1.5.1. Definitions. The *action* of a group G on a topological space X is a homomorphism

$$G \to \mathrm{Homeo}(X)$$

where $\mathrm{Homeo}(X)$ is the group of all self-homeomorphisms of X. The quotient set $X/_G$ is the set of all orbits in X and we give it the usual quotient topology. We denote by $g(x)$ the image of $x \in X$ along the homeomorphism determined by $g \in G$. The action is:

- *free* if $g(x) \neq x$ for all non-trivial $g \in G$ and all $x \in X$;
- *properly discontinuous* if any two points $x, y \in X$ have neighbourhoods U_x and U_y such that the set

$$\big\{ g \in G \mid g(U_x) \cap U_y \neq \varnothing \big\}$$

  is finite.

The relevance of these definitions is due to the following.

Proposition 1.5.1. *Let G act on a Hausdorff connected space X. The following are equivalent:*

(1) *G acts freely and properly discontinuously;*
(2) *the quotient $X/_G$ is Hausdorff and $X \to X/_G$ is a covering.*

A covering of type $X \to X/_G$ is called *regular*. A covering $X \to Y$ is regular if and only if the image of $\pi_1(X)$ in $\pi_1(Y)$ is normal, and in that case G is the quotient of the two groups. In particular every universal cover is regular. Summing up:

Corollary 1.5.2. *Every path-connected locally contractible Hausdorff topological space X is the quotient $\tilde{X}/_G$ of its universal cover by the action of some group G acting freely and properly discontinuously.*

The group G is isomorphic to $\pi_1(X)$.

Exercise 1.5.3. Let a discrete group G act on a locally compact space X. The following are equivalent:

- the action is properly discontinuous;
- for every compact $K \subset X$, the set $\{g \mid g(K) \cap K \neq \varnothing\}$ is finite;
- the map $G \times X \to X \times X$ that sends (g, x) to $(g(x), x)$ is proper.

1.5.2. Isometry group. We now want to consider the case where a group G acts by isometries on a Riemannian manifold M. In this case we have a homomorphism

$$G \to \mathsf{Isom}(M)$$

with values in the *isometry group* $\mathsf{Isom}(M)$ of M, *i.e.* the group of all isometries $f: M \to M$. The group $\mathsf{Isom}(M)$ is much smaller than $\mathsf{Homeo}(M)$.

Theorem 1.5.4 (Myers – Steenrod). *The group* $\mathsf{Isom}(M)$ *has a natural Lie group structure compatible with the compact-open topology. The map* $F: \mathsf{Isom}(M) \times M \to M \times M$ *that sends* (φ, p) *to* $(\varphi(p), p)$ *is proper.*

Corollary 1.5.5. *If M is compact then* $\mathsf{Isom}(M)$ *is compact.*

Proposition 1.5.6. *The stabilizer of any point $x \in M$ is a compact Lie subgroup of* $\mathsf{Isom}(M)$.

Proof. The stabilizer of x is closed and hence is a Lie subgroup of $\mathsf{Isom}(M)$. The isometries that fix x are determined by their orthogonal action on $T_x M$ and therefore form a compact Lie subgroup of $O(n)$. $\square$

The hard part of the Myers – Steenrod Theorem is to endow $\mathsf{Isom}(M)$ with a Lie group structure. In fact, in all the concrete cases that we will encounter, the Lie group structure of $\mathsf{Isom}(M)$ will be evident from the context, so we will not need the full strength of Theorem 1.5.4. The rest of the theorem is not difficult to prove and we can leave it as an exercise.

Exercise 1.5.7. Prove that F is proper.

We denote by $\mathsf{Isom}^+(M)$ the subgroup of $\mathsf{Isom}(M)$ consisting of all the orientation-preserving isometries. It has index one or two in $\mathsf{Isom}(M)$.

1.5.3. Discrete groups. Let M be a Riemannian manifold. The group $\mathsf{Isom}(M)$ is a Lie group, so it makes sense to consider discrete subgroups.

Proposition 1.5.8. *A group $\Gamma < \mathsf{Isom}(M)$ acts properly discontinuously on M if and only if it is discrete.*

Proof. If Γ is discrete, Theorem 1.5.4 and Exercise 1.5.3 imply that it acts properly discontinuously. Conversely, if Γ acts properly discontinuously then $e \in \Gamma$ is easily seen to be isolated, and we apply Exercise 1.4.18. $\square$

If $\Gamma < \mathsf{Isom}(M)$ is discrete and acts freely, the map $M \to M/_\Gamma$ is a covering. Moreover, the Riemannian structure projects from M to $M/_\Gamma$.

Proposition 1.5.9. *Let $\Gamma < \mathrm{Isom}(M)$ act freely and properly discontinuously on M. There is a unique Riemannian structure on the manifold $M/_\Gamma$ such that the covering $\pi \colon M \to M/_\Gamma$ is a local isometry.*

Proof. Let $U \subset M/_\Gamma$ be a well-covered set: we have $\pi^{-1}(U) = \sqcup_{i \in I} U_i$ and $\pi|_{U_i} \colon U_i \to U$ is a homeomorphism. Pick an $i \in I$ and assign to U the smooth and Riemannian structure of U_i transported along π. The resulting structure on U does not depend on i since the open sets U_i are related by isometries in Γ. We get a Riemannian structure on $M/_\Gamma$, determined by the fact that π is a local isometry. □

1.5.4. Measures. The case where a group G acts by preserving some measure instead of a Riemannian metric is also very interesting, although quite different. This situation will occur only in Chapter 8.

Let a group G act on a manifold M. A Borel measure μ on M is *G-invariant* if $\mu(S) = \mu(g(S))$ for every Borel set $S \subset M$.

If G acts on M freely and properly discontinuously, then $M \to M/_G$ is a covering and every G-invariant measure μ on M descends to a natural measure on the quotient $M/_G$ which we still indicate by μ. The measure on $M/_G$ is defined as follows: for every well-covered open set $U \subset M/_G$ we have $\pi^{-1}(U) = \sqcup_{i \in I} U_i$ and we assign to U the measure of U_i for any $i \in I$. This assignment extends to a unique Borel measure on $M/_G$ by Proposition 1.3.4.

Note that the measure on $M/_G$ is *not* the push-forward of μ, namely it is not true that $\mu(U) = \mu(\pi^{-1}(U))$.

1.6. Homology

The singular homology theory needed in this book is not very deep: all the homology groups of the manifolds that we consider are boringly determined by their fundamental group. The theory is quickly reviewed in this section, a standard introduction is Hatcher's *Algebraic Topology* [**25**].

1.6.1. Definition. Let X be a topological space and R be a ring. The cases $R = \mathbb{Z}$, $\mathbb{R}$, or $\mathbb{Z}/_{2\mathbb{Z}}$ are typically the most interesting ones.

A *singular k-simplex* is a continuous map $\alpha \colon \Delta_k \to X$ from the standard k-dimensional simplex Δ_k into X. A *k-chain* is an abstract linear combination

$$\lambda_1 \alpha_1 + \ldots + \lambda_h \alpha_h$$

of singular k-simplexes $\alpha_1, \ldots, \alpha_h$ with coefficients $\lambda_1, \ldots, \lambda_h \in R$. The set $C_k(X, R)$ of all k-chains is a R-module. There is a linear boundary map $\partial_k \colon C_k(X, R) \to C_{k-1}(X, R)$ such that $\partial_{k-1} \circ \partial_k = 0$. The *cycles* and *boundaries* are the elements of the submodules

$$Z_k(X, R) = \ker \partial_k, \quad B_k(X, R) = \mathrm{Im}\, \partial_{k+1}.$$

The k-th *homology group* is the quotient

$$H_k(X, R) = Z_k(X, R)/_{B_k(X,R)}.$$

We sometimes omit R and write $H_k(X)$ instead of $H_k(X, R)$.

By taking the dual spaces $C^k(X, R) = \mathrm{Hom}(C_k(X, R), R)$ we define analogously the *cohomology group* $H^k(X, R)$. (Co-)homology groups are also defined for pairs (X, Y) with $Y \subset X$. When $X = U \cup V$ and U, V are open we get the exact *Mayer − Vietoris sequence*:

$$\ldots \longrightarrow H_{n+1}(X) \longrightarrow H_n(U \cap V) \longrightarrow H_n(U) \oplus H_n(V) \longrightarrow H_n(X) \longrightarrow \ldots$$

(Co-)homology groups are *functorial* in the sense that continuous maps induce natural homomorphisms of groups. Homotopic maps induce the same homomorphisms.

If X is path-connected, then $H_0(X, R) = R$ and there is a canonical homomorphism

$$\pi_1(X) \longrightarrow H_1(X, \mathbb{Z}).$$

The homomorphism is surjective and its kernel is generated by the commutators of $\pi_1(X)$: in other words $H_1(X, \mathbb{Z})$ is the *abelianization* of $\pi_1(X)$. Concerning cohomology, we have

$$H^1(X, \mathbb{Z}) = \mathrm{Hom}\big(H_1(X, \mathbb{Z}), \mathbb{Z}\big) = \mathrm{Hom}(\pi_1(X), \mathbb{Z}).$$

The relation between $H^i(X, \mathbb{Z})$ and $H_i(X, \mathbb{Z})$ for $i > 1$ is not so immediate.

1.6.2. Dualities. Let M be a compact oriented connected n-manifold with (possibly empty) boundary. The abelian group $H_k(M, \mathbb{Z})$ is finitely generated and hence decomposes as

$$H_k(M, \mathbb{Z}) \cong F_k \oplus T_k$$

where $F_k = \mathbb{Z}^{b_k}$ is free and T_k is finite. The *torsion* subgroup T_k consists of all finite-order elements in $H_k(M, \mathbb{Z})$. The rank b_k of F_k is the k-th *Betti number* of M. In cohomology things change only a little:

$$H^k(M, \mathbb{Z}) \cong F_k \oplus T_{k-1}.$$

All these groups vanish when $k > n$. Even when the torsion vanishes, there is no canonical isomorphism between $H_k(M)$ and $H^k(M)$. On the other hand, the *Lefschetz duality* provides a canonical identification

$$H^k(M) = H_{n-k}(M, \partial M)$$

for any ring R. When $\partial M = \varnothing$ this is the *Poincaré duality* $H^k(M) = H_{n-k}(M)$. In particular we get

$$H_n(M, \partial M, \mathbb{Z}) = H^0(M, \mathbb{Z}) \cong \mathbb{Z}$$

and the choice of an orientation for M is equivalent to a choice of a generator $[M] \in H_n(M, \partial M, \mathbb{Z})$ called the *fundamental class* of M.

An important exact sequence for M is the following:

$$\ldots \longrightarrow H_n(M) \longrightarrow H_n(M, \partial M) \longrightarrow H_{n-1}(\partial M) \longrightarrow H_{n-1}(M) \longrightarrow \ldots$$

1.6.3. Intersection form. Let G and H be finitely generated abelian groups, seen as $\mathbb{Z}$-modules. A bilinear form

$$\omega \colon G \times H \longrightarrow \mathbb{Z}$$

is *non-degenerate* if for every infinite-order element $g \in G$ there is a $h \in H$ such that $\omega(g, h) \neq 0$. If $G = H$, we say that ω is *symmetric* (resp. *skew-symmetric*) if $\omega(g_1, g_2)$ equals $\omega(g_2, g_1)$ (resp. $-\omega(g_2, g_1)$) for all $g_1, g_2 \in G$. A skew-symmetric non-degenerate form is called *symplectic*.

Let M be a compact oriented connected n-manifold with (possibly empty) boundary. The Lefschetz duality furnishes a non-degenerate bilinear form

$$\omega \colon H_k(M, \mathbb{Z}) \times H_{n-k}(M, \partial M, \mathbb{Z}) \longrightarrow \mathbb{Z}$$

called the *intersection form*. It has the following geometric interpretation.

An oriented closed k-submanifold $S \subset M$ defines a class $[S] \in H_k(M)$ as the image of its fundamental class via the map $i_* \colon H_k(S) \to H_k(M)$ induced by the inclusion. If S has boundary and is properly embedded (that is, $\partial S = \partial M \cap S$), it defines a class $[S] \in H_k(M, \partial M)$.

Suppose two oriented submanifolds S and S' have complementary dimensions k and $n - k$ and intersect transversely: every intersection point x is isolated and has a sign ± 1, defined by comparing the orientations of $T_x S \oplus T_x S'$ and $T_x M$. The *algebraic intersection* $S \cdot S'$ of S and S' is the sum of these signs.

Theorem 1.6.1. *Let S, S' be transverse and represent two classes $[S] \in H_k(M)$ and $[S'] \in H_{n-k}(M, \partial M)$. We have*

$$\omega([S], [S']) = S \cdot S'.$$

Corollary 1.6.2. *The intersection number of two transverse oriented submanifolds of complementary dimension depends only on their homology classes.*

When M is closed and has even dimension $2n$, the central form

$$\omega \colon H_n(M, \mathbb{Z}) \times H_n(M, \mathbb{Z}) \longrightarrow \mathbb{Z}$$

is symmetric or skew-symmetric, depending on whether n is even or odd.

Everything we said about fundamental classes, Lefschetz duality, and intersection forms holds for non-orientable manifolds as well, provided that we pick $R = \mathbb{Z}/_{2\mathbb{Z}}$ and consider bilinear forms and intersections in $\mathbb{Z}/_{2\mathbb{Z}}$.

1.7. Cells and handle decompositions

Many nice topological spaces can be constructed iteratively starting from finitely many points and then attaching some discs of increasing dimension called *cells*. When the topological space is a differentiable manifold one typically thickens the cells to *handles*.

Cells and handles are beautifully introduced by Hatcher [**25**] and Kosinksi [**34**], respectively.

1.7.1. Cell complexes. A *finite cell complex* of dimension n (briefly, a n-complex) is a topological space obtained iteratively in the following manner:

- a 0-complex X^0 is a finite set of points,
- a n-complex X^n is obtained from a $(n-1)$-complex X^{n-1} by attaching finitely many n-cells, that is copies of D^n glued along continuous maps $\varphi \colon \partial D^n \to X^{n-1}$.

The closed subset $X^k \subset X^n$ is the *k-skeleton* of X^n, for all $k < n$.

Proposition 1.7.1. *Let X be a n-complex. The inclusion map $X^k \hookrightarrow X$ induces an isomorphism $\pi_j(X^k) \to \pi_j(X)$ for all $j < k$.*

Proof. By transversality, maps $S^j \to X$ and homotopies between them can be homotoped away from cells of dimension $\geqslant j + 2$. □

In particular, the space X is connected if and only if X^1 is, and its fundamental group is captured by X^2.

Theorem 1.7.2. *Every differentiable compact n-manifold may be realised topologically as a finite n-complex.*

A presentation for the fundamental group of a cell complex X can be constructed as follows. If $x_0 \in X^0$, we fix a maximal tree $T \subset X^1$ containing x_0 and equip the k arcs in $X^1 \setminus T$ with some arbitrary orientations. These arcs determine some generators $g_1, \ldots, g_k \in \pi_1(X, x_0)$. The boundary of a 2-cell makes a circular path in X^1: every time it crosses an arc g_i in some direction (entering from one side and exiting from the other) we write the corresponding letter $g_i^{\pm 1}$ and get a word. The s two-cells produce s word relations. We have constructed a presentation for $\pi_1(X)$.

1.7.2. Euler characteristic. The *Euler characteristic* of a n-complex X is the integer

$$\chi(X) = \sum_{i=0}^{n} (-1)^i C_i$$

where C_i is the number of i-cells in X. It is also equal to

$$\chi(X) = \sum_{i=0}^{n} (-1)^i b_i(X)$$

where $b_i(X)$ is the i-th Betti number of X. Therefore $\chi(X)$ is a number which depends only on the topology of X, that can be easily calculated from any cell decomposition of X.

Proposition 1.7.3. *If $\tilde{X} \to X$ is a degree-d covering of finite complexes then $\chi(\tilde{X}) = d \cdot \chi(X)$.*

Proof. A k-cell in X is simply-connected and hence lifts to d distinct k-cells in $\tilde{X}$. A cell decomposition of X thus induces one of $\tilde{X}$ where the numbers C_i are all multiplied by d. □

It is much harder to control homology under coverings. The Euler characteristic of a closed manifold measures the obstruction of constructing a nowhere-vanishing vector field.

Theorem 1.7.4. *A closed differentiable manifold M has a nowhere-vanishing vector field if and only if $\chi(M) = 0$.*

1.7.3. Aspherical cell-complexes. A finite cell complex is locally contractible and hence has a universal cover $\tilde{X}$. If $\tilde{X}$ is contractible the complex X is called *aspherical*.

Example 1.7.5. Every closed manifold M covered by $\mathbb{R}^n$ is aspherical.

By a theorem of Whitehead, a finite cell complex X is aspherical if and only if all its higher homotopy groups $\pi_i(X)$ with $i \geqslant 2$ vanish.

The following theorem says that maps to aspherical spaces are determined (up to homotopy) by homomorphisms between fundamental groups.

Theorem 1.7.6. *Let $(X, x_0), (Y, y_0)$ be pointed connected finite cellular complexes. If Y is aspherical, every homomorphism $\pi_1(X, x_0) \to \pi_1(Y, y_0)$ is induced by a continuous map $(X, x_0) \to (Y, y_0)$, unique up to homotopy.*

Proof. We construct a continuous map $f \colon X^k \to Y^k$ iteratively on the k-skeleta, starting from $k = 1$.

Let T be a maximal tree in X^1. The oriented 1-cells $g_1, \ldots, g_k$ in $X^1 \setminus T$ define generators in $\pi_1(X, x_0)$. We define $f \colon X^1 \to Y^1$ by sending T to y_0 and each g_i to any loop in Y representing the image of g_i along the given homomorphism $\pi_1(X, x_0) \to \pi_1(Y, y_0)$.

The map f sends the boundary of each 2-cell to a homotopically trivial loop in Y and hence extends to a map $f \colon X^2 \to Y^2$. Since Y is aspherical, the higher homotopy groups $\pi_i(Y)$ with $i \geqslant 2$ vanish and hence f extends to a map $f \colon X^k \to Y^k$ iteratively for all $k \geqslant 3$.

We prove that f is unique up to homotopy. Let f' be a map that realises the given homomorphism on fundamental groups. We can construct a homotopy between f and f' iteratively on X^k as follows.

For $k = 1$, we can suppose that both f and f' send T to y_0. By hypothesis they send the generators g_i to homotopic loops, hence we can homotope f' to f on X^1. For $k \geqslant 2$, the maps f and f' on each k-cell are homotopic because they glue to a map $S^k \to Y$, which is null-homotopic because $\pi_k(Y)$ is trivial. $\square$

Corollary 1.7.7. *Let X and Y be connected finite aspherical complexes. Every isomorphism $\pi_1(X) \to \pi_1(Y)$ is realised by a homotopic equivalence $X \to Y$, unique up to homotopy.*

In particular the homotopy type of an aspherical manifold is fully determined by its fundamental group.

Corollary 1.7.8. *Two aspherical closed manifolds of distinct dimension have non-isomorphic fundamental groups.*

Proof. Two closed manifolds of different dimension cannot be homo-topically equivalent because they have non-isomorphic homology groups. ∎

We cite for completeness this result, although we will never use it.

Theorem 1.7.9 (Cartan–Hadamard). *A complete Riemannian mani-fold M with sectional curvature everywhere $\leqslant 0$ has a universal covering diffeomorphic to $\mathbb{R}^n$ and is hence aspherical.*

Sketch of the proof. Pick a point $x \in M$. Since M is complete, the exponential map $\exp_x \colon T_x M \to M$ is defined on $T_x M$. The fact that the sectional curvatures are $\leqslant 0$ imply that $(d\exp_x)_y$ is invertible for any $y \in T_x$ and $\exp_x$ is a covering. ∎

At a single point in this book we will need the following theorem.

Theorem 1.7.10. *The fundamental group of an aspherical manifold M has no torsion.*

Sketch of the proof. Up to passing to a cover, it suffices to consider the case $\pi_1(M) \cong \mathbb{Z}/_{n\mathbb{Z}}$ for some $n \geqslant 2$. This case is excluded because the cohomology of M is isomorphic to the (suitably defined) cohomology of $\mathbb{Z}/_{n\mathbb{Z}}$, which has however infinite dimension. ∎

1.7.4. Gluing portion of boundaries. Every compact differentiable n-manifold M can be obtained topologically as a finite complex. The finite complex structure is however not designed to describe the smooth structure of M, and for that purpose it is better to replace k-cells with some thickened objects called *k-handles*. These handles are n-discs glued iteratively along portions of their boundaries. Before describing them, we briefly explain how smooth manifolds can be glued along portions of their boundaries.

Let M and N be two n-manifolds with boundary and $X \subset \partial M, Y \subset \partial N$ be two compact $(n-1)$-submanifolds with boundary. A diffeomorphism $\varphi \colon X \to Y$ defines a topological space

$$M \cup_\varphi N$$

that may be promoted to a differentiable manifold: it suffices to use a collar as in Section 1.1.12, see Figure 1.5. The glued manifold depends (up to diffeomorphism) only on the isotopy class of φ.

1.7.5. Handles. Let M be a (possibly empty or disconnected) n-manifold with boundary and $0 \leqslant k \leqslant n$. A *k-handle* is a manifold $D^k \times D^{n-k}$ attached to M along some diffeomorphism $\varphi \colon \partial D^k \times D^{n-k} \to Y \subset \partial M$, hence producing a new manifold M'.

For instance, a 0-handle is a $D^0 \times D^n = D^n$ attached to nothing, since $\partial D^0 = \varnothing$. This means that attaching a 0-handle to M consists of adding
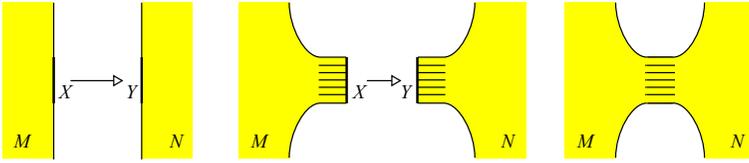
Figure 1.5. To glue two smooth manifolds along portions of their boundaries (left) we pick a collar for these portions (centre), then we remove X, Y and identify the interiors of the collars (right).
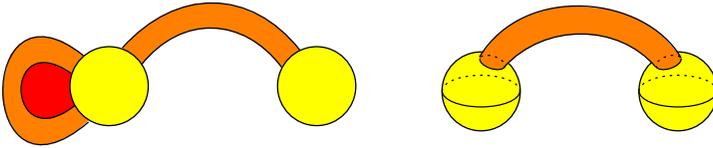
Figure 1.6. Two 0-handles (yellow), two 1-handles (orange), one 2-handle (red) in dimension 2 (left). Two 0-handles (yellow), a 1-handle (orange) in dimension 3 (right).

a disjoint disc D^n to it. For instance, by attaching a 0-handle to the empty set we create a disc D^n out of nothing.

A 1-handle is a $D^1 \times D^{n-1}$ attached along $\partial D^1 \times D^{n-1} = S^0 \times D^{n-1}$, that is two copies of D^{n-1}. Some examples in dimension $n = 2, 3$ are shown in Figure 1.6. A 2-handle is a $D^2 \times D^{n-2}$ attached along $\partial D^2 \times D^{n-2} = S^1 \times D^{n-2}$. When $n = 2$ this is a disc attached along its boundary, see Figure 1.6-(left).

1.7.6. Handle decompositions. A sequence of handle attachments

$$\varnothing \rightsquigarrow M_1 \rightsquigarrow \ldots \rightsquigarrow M_k = M$$

starting from the empty set and producing a compact manifold M with (possibly empty) boundary is called a *handle decomposition* for M. Using Morse theory one proves the following.

Theorem 1.7.11. *Every compact manifold (possibly with boundary) can be obtained from a handle decomposition.*

Example 1.7.12. The disc D^n has an obvious handle decomposition consisting of a single 0-handle. It also has more complicate handle decompositions, as Figure 1.6 shows.

By transversality, handles may always be reordered so that the lower index handles are attached first, and handles of the same index are attached simultaneously. So we can think of a decomposition as the appearing of
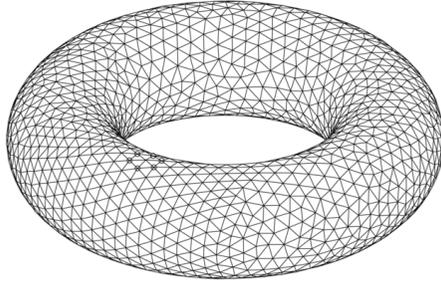
Figure 1.7. A triangulated torus.

some 0-handles, then the simultaneous attaching of some 1-handles, then of some 2-handles, and so on.

A handle decomposition of a closed manifold may be turned upside down, by reversing all arrows and interpreting every k-handle as a $(n-k)$-handle.

Proposition 1.7.13. *Every compact connected manifold M has a handle decomposition with one 0-handle and at most one n-handle.*

If M has a handle decomposition with n_i handles of index i then

$$\chi(M) = \sum_{i=0}^{n} (-1)^i n_i.$$

1.7.7. Triangulations. Instead of decomposing a manifold into handles, one may decide to decompose it into simplexes as in Figure 1.7. One such decomposition is called a triangulation: we now give a formal definition.

An (abstract and finite) *simplicial complex* is a finite set S of *vertices* together with a set X of subsets of S, called *faces*, which contains all singletons and such that every subset of a face is also a face. A face of cardinality $k+1$ has *dimension k*.

A simplicial complex X has a topological realisation $|X|$, in which every face $F \in X$ of dimension k transforms into a k-simplex with vertices in F (we get a cell complex). The complex X is *purely n-dimensional* if every face is contained in a face of dimension n. This implies that $|X|$ is the union of its n-simplexes.

Let now M be a differentiable n-manifold. A *smooth k-simplex* in M is a subset $\Delta \subset M$ diffeomorphic to a k-simplex in $\mathbb{R}^n$ (we mean that there is a diffeomorphism between open neighbourhoods of Δ and of the simplex sending the first to the second). A *smooth triangulation* of M is a pure n-dimensional simplicial complex X together with a homeomorphism

between $|X|$ and M that transforms every k-simplex in X into a smooth k-simplex in M.

Theorem 1.7.14. *Every compact manifold with (possibly empty) boundary has a smooth triangulation.*

Handle decompositions are often more flexible and efficient than triangulations, but triangulations have a more combinatorial nature and can for instance be handled by a computer. Both constructions are widely used.

On a simplicial complex X, the *star* of a vertex v is the set of all simplexes containing v plus all their sub-simplexes; the *link* of v is the subset of the star consisting of the simplexes that do not contain v. Topologically, the star of v is a closed neighbourhood of v homeomorphic to a cone on its link.

1.7.8. Hypersurfaces and cohomology. The techniques introduced in this chapter furnish another characterisation of the first cohomology group of a manifold. Let $[X, Y]$ denote the homotopy classes of continuous maps $X \to Y$.

Proposition 1.7.15. *Let X be a cell complex. There is a canonical identification*
$$H^1(X; \mathbb{Z}) = [X, S^1].$$

Proof. We know that $H^1(X; \mathbb{Z}) = \mathrm{Hom}(\pi_1(X), \mathbb{Z})$. The circle S^1 is aspherical with $\pi_1(S^1) = \mathbb{Z}$, so by Theorem 1.7.6 every homomorphism $\pi_1(X) \to \pi_1(S^1)$ is realised by a map $X \to S^1$, unique up to homotopy. □

We have noted that an oriented k-submanifold $S \subset M$ determines a class $[S] \in H_k(M, \mathbb{Z})$. It is natural to ask whether every homology class may be represented by an oriented submanifold: this is false in general, but it is true in codimension one.

Proposition 1.7.16. *Let M be a compact oriented n-manifold with (possibly empty) boundary. Every class in $H_{n-1}(M, \partial M; \mathbb{Z})$ is represented by an oriented properly embedded hypersurface $S \subset M$.*

Proof. We have
$$H_{n-1}(M, \partial M; \mathbb{Z}) = H^1(M; \mathbb{Z}) = \mathrm{Hom}(\pi_1(M), \mathbb{Z}) = [M, S^1].$$
Every map $M \to S^1$ is homotopic to a smooth map, hence each class $\alpha \in H^1(M, \mathbb{Z})$ is represented by a smooth map $f \colon M \to S^1$. By Sard's lemma there is a regular value $x \in S^1$, whose counter-image $S = f^{-1}(x)$ is a hypersurface, transversely oriented by f and hence oriented (because M is).

To prove that $[S] = \alpha$ we verify that $[S]$ and α act on $\pi_1(M)$ in the same way. Pick a loop $\gamma \in \pi_1(M)$ transverse to S. The number $\alpha(\gamma)$ is the degree of $f \circ \gamma$, that is the number of times $f \circ \gamma$ crosses x counted

with signs, which equals the number of times γ crosses S counted with signs. $\square$

Proposition 1.7.17. *An oriented connected properly embedded hypersurface $S \subset M$ is separating if and only if $[S] \in H_{n-1}(M, \partial M; \mathbb{Z})$ is trivial.*

Proof. The surface S is separating if and only if $S \cdot \alpha = 0$ for every loop α transverse to S (exercise). Since $\omega([S], [\alpha]) = S \cdot \alpha$ and ω is non-degenerate (see Section 1.6.3), this holds precisely when $[S] = 0$. $\square$

1.7.9. Topological discs. Finally, we mention a purely topological theorem that will be used only at one point in this book, in Chapter 8.

Theorem 1.7.18. *Let M be a compact topological manifold, whose interior and boundary are homeomorphic respectively to B^n and S^{n-1}. Then M is homeomorphic to D^n.*

Proof. We use two important theorems on topological manifolds, whose proofs can be found in the first chapter of the book *Topological embeddings* of Rushing [**51**].

As in every compact topological manifold, the boundary ∂M has a topological collar by a theorem of Brown (see Theorem 1.7.4 in that book). The interior boundary sphere of the collar is contained in $\text{int}(M) \cong B^n$ and hence bounds a closed disc by the Generalised Schönflies Theorem (see Theorem 1.8.2 there). The manifold M is obtained by collaring a closed disc and is hence a closed disc. $\square$

Hyperbolic space

In every dimension $n \geqslant 2$ there exists a unique simply connected complete Riemannian manifold with constant sectional curvature 1, 0, or -1. These are the sphere S^n, the Euclidean space $\mathbb{R}^n$, and the hyperbolic space $\mathbb{H}^n$.

These manifolds are the three most important spaces in Riemannian geometry. We introduce in this chapter the least familiar and the most interesting of the three: hyperbolic space.

2.1. The models of hyperbolic space

In contrast with S^n and $\mathbb{R}^n$, the hyperbolic space $\mathbb{H}^n$ may be described in various different ways, no-one of which is prevalent in the literature. Each description is a *model* for $\mathbb{H}^n$. We first introduce the hyperboloid model, which has a more algebraic flavour, and then we turn to the disc and half-space models that are somehow more geometric (and easier to visualise in dimensions $n = 2$ and 3).

2.1.1. Hyperboloid. The sphere S^n is the set of all points with norm 1 in $\mathbb{R}^{n+1}$, equipped with the Euclidean scalar product. Analogously, we may define $\mathbb{H}^n$ as the set of all points of norm -1 in $\mathbb{R}^{n+1}$, equipped with the usual *Lorentzian* scalar product. This set forms a hyperboloid with two sheets, and we choose one.

Definition 2.1.1. The *Lorentzian scalar product* on $\mathbb{R}^{n+1}$ is:

$$\langle x, y \rangle = \sum_{i=1}^{n} x_i y_i - x_{n+1} y_{n+1}.$$

It has signature $(n, 1)$. A vector $x \in \mathbb{R}^{n+1}$ is *time-like*, *light-like*, or *space-like* if $\langle x, x \rangle$ is negative, null, or positive respectively. The *hyperboloid model* I^n is defined as follows:

$$I^n = \left\{ x \in \mathbb{R}^{n+1} \mid \langle x, x \rangle = -1, \ x_{n+1} > 0 \right\}.$$

The set of points x with $\langle x, x \rangle = -1$ is a *hyperboloid with two sheets*, and I^n is the connected component (sheet) with $x_{n+1} > 0$. Let us prove a general fact. For us, a scalar product is a real non-degenerate symmetric bilinear form.
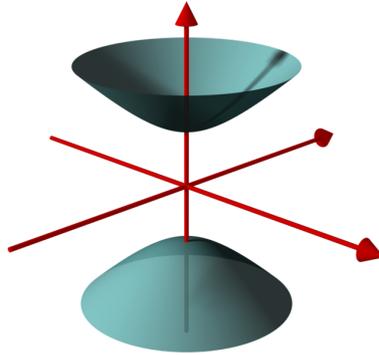
Figure 2.1. The hyperboloid with two sheets defined by the equation $\langle x, x \rangle = -1$. The model I^n for $\mathbb{H}^n$ is the upper sheet.

Proposition 2.1.2. *Let $\langle , \rangle$ be a scalar product on $\mathbb{R}^{n+1}$. The function $f \colon \mathbb{R}^{n+1} \to \mathbb{R}$ given by*

$$f(x) = \langle x, x \rangle$$

is everywhere smooth and has differential

$$df_x(y) = 2\langle x, y \rangle.$$

Proof. The following equality holds:

$$\langle x + y, x + y \rangle = \langle x, x \rangle + 2\langle x, y \rangle + \langle y, y \rangle.$$

The component $\langle x, y \rangle$ is linear in y while $\langle y, y \rangle$ is quadratic.   □

Corollary 2.1.3. *The hyperboloid I^n is a Riemannian manifold.*

Proof. The hyperboloid is the set of points with $f(x) = \langle x, x \rangle = -1$. For all $x \in I^n$ the differential $y \mapsto 2\langle x, y \rangle$ is surjective and hence the hyperboloid is a differentiable submanifold of codimension 1.

The tangent space $T_x I^n$ at $x \in I^n$ is the hyperplane

$$T_x = \ker df_x = \left\{ y \mid \langle x, y \rangle = 0 \right\} = x^{\perp}$$

orthogonal to x in the Lorentzian scalar product. Since x is time-like, the restriction of the Lorentzian scalar product to $x^{\perp}$ is positive definite and hence defines a metric tensor on I^n.   □

The hyperboloid I^n is a model for hyperbolic space $\mathbb{H}^n$. We will prove that it is simply connected, complete, and has constant curvature -1.

2.1.2. Isometries of the hyperboloid. The isometries of I^n are easily classified using linear algebra.

Let $O(n, 1)$ be the group of linear isomorphisms f of $\mathbb{R}^{n+1}$ that preserve the Lorentzian scalar product, *i.e.* such that $\langle v, w \rangle = \langle f(v), f(w) \rangle$ for any $v, w \in \mathbb{R}^n$. An element in $O(n, 1)$ preserves the hyperboloid with two sheets, and the elements preserving the upper sheet I^n form a subgroup of index two in $O(n, 1)$ that we indicate with $O^+(n, 1)$.

Proposition 2.1.4. *The following equality holds:*

$$\text{Isom}(I^n) = O^+(n, 1).$$

Proof. Pick $f \in O^+(n, 1)$. If $x \in I^n$ then $f(x) \in I^n$ and f sends $x^\perp$ to $f(x)^\perp$ isometrically, hence $f \in \text{Isom}(I^n)$. Therefore $O^+(n, 1) \subseteq \text{Isom}(I^n)$.

To prove the converse inclusion we show that for every pair $x, y \in I^n$ and every linear isometry $g \colon x^\perp \to y^\perp$ there is an element $f \in O^+(n, 1)$ such that $f(x) = y$ and $f|_{x^\perp} = g$. Since isometries are determined by their first-order behaviour at a point x, this implies $\text{Isom}(I^n) \subseteq O^+(n, 1)$.

Via elementary linear algebra we prove that $O^+(n, 1)$ acts transitively on I^n and hence we may suppose that $x = y = (0, \dots, 0, 1)$. Now $x^\perp = y^\perp$ is the horizontal hyperplane and $g \in O(n)$. To define f simply take

$$f = \begin{pmatrix} g & 0 \\ 0 & 1 \end{pmatrix}.$$

The proof is complete.   □

The isometry groups of S^n and $\mathbb{R}^n$ are described analogously:

Proposition 2.1.5. *The following equalities hold:*

$$\text{Isom}(S^n) = O(n + 1),$$
$$\text{Isom}(\mathbb{R}^n) = \big\{ x \mapsto Ax + b \mid A \in O(n), b \in \mathbb{R}^n \big\}.$$

Proof. The proof is analogous to the one above.   □

We have also proved the following fact. A *frame* at a point p in a Riemannian manifold M is an orthonormal basis for $T_p M$.

Corollary 2.1.6. *Let $M = S^n$, $\mathbb{R}^n$, or $\mathbb{H}^n$. Given two points $p, q \in M$ and two frames at p and q, there is a unique isometry that carries the first frame to the second.*

2.1.3. Subspaces. We introduce the following natural objects.

Definition 2.1.7. A k-dimensional *subspace* of $\mathbb{R}^n$, S^n, I^n is:
- an affine k-dimensional space in $\mathbb{R}^n$,
- the intersection of a $(k+1)$-dimensional vector subspace of $\mathbb{R}^{n+1}$ with S^n,
- the intersection of a $(k+1)$-dimensional vector subspace of $\mathbb{R}^{n+1}$ with I^n, when it is non-empty.

Remark 2.1.8. Elementary linear algebra shows that the following conditions are equivalent for any $(k + 1)$-dimensional vector subspace $W \subset \mathbb{R}^{n+1}$:

(1) $W \cap I^n \neq \varnothing$,
(2) W contains at least a time-like vector,
(3) the signature of $\langle, \rangle|_W$ is $(k, 1)$.

A k-subspace in $\mathbb{R}^n, S^n, \mathbb{H}^n$ is itself isometric to $\mathbb{R}^k, S^k, \mathbb{H}^k$. The non-empty intersection of two subspaces is always a subspace. An isometry of $\mathbb{R}^n, S^n, \mathbb{H}^n$ sends k-subspaces to k-subspaces.

Exercise 2.1.9. Let S be a k-subspace in $\mathbb{R}^n, S^n$, or $\mathbb{H}^n$ and $p \in S$ a point. There is a unique $(n - k)$-subspace S' intersecting S orthogonally in p.

2.1.4. Reflections. We now introduce a basic kind of isometry. The *reflection* r_S along a subspace S in I^n is an isometry of I^n defined as follows. By definition $S = I^n \cap W$ with $\langle, \rangle|_W$ non-degenerate, hence $\mathbb{R}^{n+1} = W \oplus W^\perp$ and we set $r_S|_W = \text{id}$ and $r_S|_{W^\perp} = -\text{id}$. Analogous definitions work for subspaces of S^n and $\mathbb{R}^n$.

Exercise 2.1.10. The reflection r_S has fixed set S and preserves all the subspaces orthogonal to S. It is orientation-preserving if and only if S has even codimension.

Proposition 2.1.11. *Reflections along hyperplanes generate the isometry groups of S^n, $\mathbb{R}^n$, and $\mathbb{H}^n$.*

Proof. It is a standard linear algebra fact that orthogonal reflections along vector hyperspaces generate $O(n)$. This proves the case S^n and shows that reflections generate the stabiliser of any point in $\mathbb{R}^n$ and $\mathbb{H}^n$. To conclude it suffices to check that reflections act transitively on points: to send x to y, reflect along the hyperplane orthogonal to the segment connecting x to y in its midpoint. $\qquad\square$

2.1.5. Lines. A 1-subspace is a *line*. We show that lines and geodesics are the same thing. Recall the hyperbolic trigonometric functions:

$$\sinh(t) = \frac{e^t - e^{-t}}{2}, \quad \cosh(t) = \frac{e^t + e^{-t}}{2}.$$

Proposition 2.1.12. *A non-trivial complete geodesic in S^n, $\mathbb{R}^n$, or $\mathbb{H}^n$ is a line run at constant speed. Concretely, let $p \in M$ be a point and $v \in T_p M$ a unit vector. The geodesic γ exiting from p with velocity v is:*

- $\gamma(t) = \cos(t) \cdot p + \sin(t) \cdot v$ *if* $M = S^n$,
- $\gamma(t) = p + tv$ *if* $M = \mathbb{R}^n$,
- $\gamma(t) = \cosh(t) \cdot p + \sinh(t) \cdot v$ *if* $M = I^n$.

Proof. Let $p \in I^n$ be a point, $v \in T_p M$ a unit vector, and γ the geodesic exiting from p with velocity v. The plane $W \subset \mathbb{R}^{n+1}$ generated

by p and v intersects I^n into a line l containing p and tangent to v. To prove that l is the support of γ we use a symmetry argument: the reflection r_l fixes p and v and hence γ, therefore γ is forced to be contained in its fixed locus, which is l. This shows that non-trivial geodesics are lines run at constant speed.

We now consider the curve $\alpha(t) = \cosh(t) \cdot p + \sinh(t) \cdot v$. We have $\alpha(0) = p$ and $\alpha'(0) = v$. It remains to prove that α parametrizes l with unit speed, and from this we deduce that $\gamma = \alpha$. We note that

$$\langle \alpha(t), \alpha(t) \rangle = \cosh^2(t) \langle p, p \rangle + 2 \cosh(t) \sinh(t) \langle p, v \rangle + \sinh^2(t) \langle v, v \rangle$$
$$= -\cosh^2(t) + \sinh^2(t) = -1.$$

Therefore α parametrizes l. Its velocity is

$$\alpha'(t) = \cosh'(t) \cdot p + \sinh'(t) \cdot v = \sinh(t) \cdot p + \cosh(t) \cdot v$$

whose squared norm is $-\sinh^2(t) + \cosh^2(t) = 1$. Therefore $\gamma = \alpha$. The proofs for S^n and $\mathbb{R}^n$ are analogous. $\square$

Corollary 2.1.13. *The spaces S^n, $\mathbb{R}^n$, and $\mathbb{H}^n$ are complete.*

Proof. The previous proposition shows that geodesics are defined on $\mathbb{R}$, hence the space is complete by the Hopf − Rinow Theorem 1.2.8. $\square$

It is easy to show that two points in $\mathbb{H}^n$ are contained in a unique line.

Remark 2.1.14. Euclid's fifth postulate holds only in $\mathbb{R}^2$. Given a line r and a point $P \notin r$, there is exactly one line passing through P and disjoint from r in $\mathbb{R}^2$, there is no-one in S^2, and there are infinitely many in $\mathbb{H}^2$.

We can easily calculate the distance between two points.

Proposition 2.1.15. *Let $p, q \in M$ be two points. We have*
- $\cos(d(p, q)) = \langle p, q \rangle$ *if $M = S^n$,*
- $\cosh(d(p, q)) = -\langle p, q \rangle$ *if $M = I^n$.*

Proof. Let γ be a geodesic that goes from p to q at unit speed. We have $\gamma(0) = p$ and $\gamma(t) = q$ with $t = d(p, q)$. Set $v = \gamma'(0)$ and use Proposition 2.1.12 to conclude. $\square$

2.1.6. The Poincaré disc. We introduce two models of $\mathbb{H}^n$, the *disc* and the *half-space*, that are easier to visualise especially in the dimensions $n = 2, 3$ we are interested in. The first model is the *Poincaré disc*

$$D^n = \{x \in \mathbb{R}^n \mid \|x\| < 1\}.$$

The metric tensor on D^n is obviously not the Euclidean one of $\mathbb{R}^n$, but instead is induced by a particular diffeomorphism between I^n and D^n that we construct now. We identify $\mathbb{R}^n$ with the horizontal hyperplane $x_{n+1} = 0$ in $\mathbb{R}^{n+1}$ and note that the projection towards $P = (0, \ldots, 0, -1)$
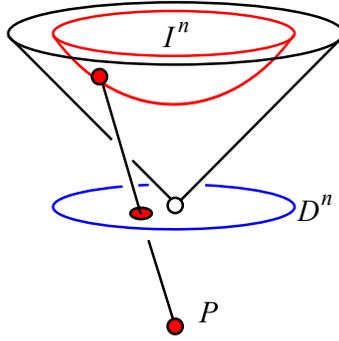
Figure 2.2. The projection towards $P = (0, \ldots, 0, -1)$ induces a bijection between the hyperboloid model I^n and the disc model D^n.

depicted in Figure 2.2 induces a bijection between I^n and the horizontal disc $D^n \subset \mathbb{R}^n$. The projection p may be written as:

$$p(x_1, \ldots, x_{n+1}) = \frac{(x_1, \ldots, x_n)}{x_{n+1} + 1}$$

and is indeed a diffeomorphism $p \colon I^n \to D^n$ that transports the metric tensor on I^n to some metric tensor g on D^n.

Proposition 2.1.16. *The metric tensor g at $x \in D^n$ is:*

$$g_x = \left(\frac{2}{1 - \|x\|^2} \right)^2 \cdot g_x^E$$

where g^E is the Euclidean metric tensor on $D^n \subset \mathbb{R}^n$.

Proof. The inverse $q \colon D^n \to I^n$ of p is:

$$q(x) = \left(\frac{2x_1}{1 - \|x\|^2}, \ldots, \frac{2x_n}{1 - \|x\|^2}, \frac{1 + \|x\|^2}{1 - \|x\|^2} \right).$$

Pick $x \in D^n$. Rotations around the x_{n+1} axis are isometries of I^n and commute with p, therefore they are isometries of (D^n, g) too. Up to rotating we may take $x = (x_1, 0, \ldots, 0)$ and find

$$dq_x = \frac{2}{1 - x_1^2} \cdot \begin{pmatrix} \frac{1 + x_1^2}{1 - x_1^2} & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \\ 2\frac{x_1}{1 - x_1^2} & 0 & \cdots & 0 \end{pmatrix}.$$

The column vectors form an orthonormal basis of $T_{q(x)} I^n$. Hence dq_x stretches all vectors of a constant $\frac{2}{1 - x_1^2}$. Therefore $g_x = \frac{4}{(1 - x_1^2)^2} g_x^E$. $\qquad \square$

Figure 2.3. Three lines that determine a hyperbolic triangle
in the Poincaré disc. The inner angles α, β, and γ coincide
with the Euclidean ones, and we have $\alpha + \beta + \gamma < \pi$.

The Poincaré disc is a *conformal* model of $\mathbb{H}^n$: it is a model where
the metric differs from the Euclidean metric only by multiplication by a
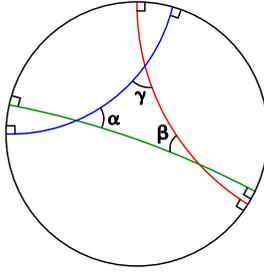positive scalar $\left(\frac{2}{1-\|x\|^2}\right)^2$ that depends smoothly on x. We note that the
scalar tends to infinity when x tends to ∂D^n. On a conformal model the
lengths of vectors are different from the Euclidean lengths, but the angles
formed by two adjacent vectors coincide with the Euclidean ones. Shortly:
lengths are distorted but angles are preserved.

Let us see how we can easily visualise k-subspaces in the disc model.

Proposition 2.1.17. *The k-subspaces in D^n are the intersections of*
D^n *with k-spheres and k-planes of $\mathbb{R}^n$ orthogonal to ∂D^n.*

Proof. Since every k-subspace is an intersection of hyperplanes, we
easily restrict to the case $k = n - 1$. A hyperplane in I^n is $I^n \cap v^\perp$ for some
space-like vector v. If v is horizontal (*i.e.* its last coordinate is zero) then
$v^\perp$ is vertical and $p(I^n \cap v^\perp) = D^n \cap v^\perp$, a hyperplane orthogonal to ∂D^n.

If v is not horizontal, up to rescaling and rotating around x_{n+1} we may
suppose $v = (\alpha, 0, \ldots, 0, 1)$ with $\alpha > 1$. The hyperplane is
$$\left\{x_1^2 + \ldots + x_n^2 - x_{n+1}^2 = -1\right\} \cap \left\{x_{n+1} = \alpha x_1\right\}.$$
On the other hand the sphere in $\mathbb{R}^n$ of centre $(\alpha, 0, \ldots, 0)$ and radius
$\sqrt{\alpha^2 - 1}$ is orthogonal to ∂D^n and is described by the equation
$$\left\{(y_1 - \alpha)^2 + y_2^2 + \ldots + y_n^2 = \alpha^2 - 1\right\} = \left\{y_1^2 + \ldots + y_n^2 - 2\alpha y_1 = -1\right\}$$
which is equivalent to $\|y\|^2 = -1 + 2\alpha y_1$. If $y = p(x)$ the relations
$$y_1 = \frac{x_1}{x_{n+1} + 1}, \quad \|y\|^2 = \frac{x_{n+1} - 1}{x_{n+1} + 1}$$
transform the latter equation in $x_{n+1} = \alpha x_1$.   □

Three lines in D^2 delimiting a hyperbolic triangle are drawn in Figure
2.3. Since the disc is a conformal model, the inner angles α, β, γ are the
ones we see on the picture. In particular we verify easily that $\alpha + \beta + \gamma < \pi$.
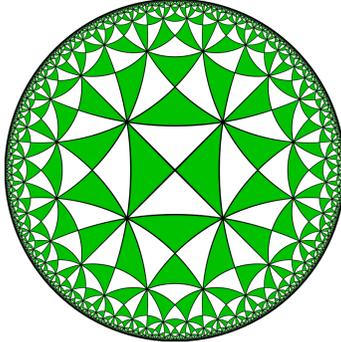
Figure 2.4. A *tessellation* of S^2, $\mathbb{R}^2$ o $\mathbb{H}^2$ is a subdivision of
the plane into polygons. The tessellation of $\mathbb{H}^2$ shown here
is obtained by drawing infinitely many lines in the plane. The
triangles have inner angles $\frac{\pi}{2}, \frac{\pi}{5}, \frac{\pi}{5}$ and are all isometric.

Exercise 2.1.18. For any triple of positive angles α, β, γ with $\alpha + \beta + \gamma < \pi$ there is a triangle with inner angles α, β, γ. This triangle is unique up to isometry.

2.1.7. Inversions. To construct our second conformal model we need to introduce a geometric transformation called inversion.

Definition 2.1.19. Let $S = S(x_0, r)$ be the sphere in $\mathbb{R}^n$ centred in x_0 and with radius r. The *inversion* along S is the map $\varphi \colon \mathbb{R}^n \setminus \{x_0\} \to \mathbb{R}^n \setminus \{x_0\}$ defined as follows:

$$\varphi(x) = x_0 + r^2 \frac{x - x_0}{\|x - x_0\|^2}.$$

The map may be extended continuously on the whole sphere S^n, identified with $\mathbb{R}^n \cup \{\infty\}$ through the stereographic projection, by setting $\varphi(x_0) = \infty$ and $\varphi(\infty) = x_0$. A geometric description is shown in Figure 2.5.

We have already talked about conformal models. More generally, a diffeomorphism $f \colon M \to N$ between two oriented Riemannian manifolds is *conformal* (respectively, *anticonformal*) if for any $p \in M$ the differential df_p is the product of a scalar dilation $\lambda_p > 0$ and an isometry that preserves (respectively, inverts) the orientation. The scalar λ_p depends smoothly on p. A conformal map preserves the angle between two tangent vectors but multiplies their lengths by λ_p.

Exercise 2.1.20. The stereographic projection is conformal.

Proposition 2.1.21. *The following hold:*
   (1) *every inversion is a smooth and anticonformal map;*
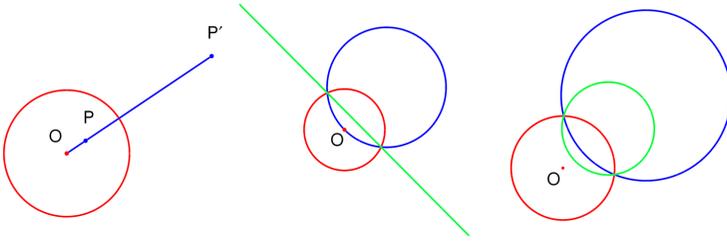
Figure 2.5. The inversion trough a sphere of centre O and radius r moves P to P' so that $OP \times OP' = r^2$ (left). It transforms a k-sphere S (blue) into a k-plane (green) if $O \in S$ (centre) or into a k-sphere (green) if $O \notin S$ (right).

(2) *inversions send k-spheres and k-planes to k-spheres and k-planes.*

Proof. Up to conjugating with translations and dilations we may suppose $x_0 = 0$ and $r = 1$. The inversion is $\varphi(x) = \frac{x}{\|x\|^2}$ and we show that $d\varphi_x$ is $\frac{1}{\|x\|^2}$ times a reflection with respect to the hyperplane orthogonal to x. We may suppose $x = (x_1, 0, \ldots, 0)$ and calculate the partial derivatives:

$$\varphi(x_1, \ldots, x_n) = \frac{(x_1, \ldots, x_n)}{\|x\|^2}, \qquad \frac{\partial \varphi_i}{\partial x_j} = \frac{\delta_{ij}\|x\|^2 - 2x_i x_j}{\|x\|^4}.$$

The partial derivatives at $x = (x_1, 0, \ldots, 0)$ are

$$\frac{\partial \varphi_1}{\partial x_1} = -\frac{1}{x_1^2}, \qquad \frac{\partial \varphi_i}{\partial x_i} = \frac{1}{x_1^2}, \qquad \frac{\partial \varphi_j}{\partial x_k} = 0$$

for all $i > 1$ and $j \neq k$. This proves our assertion.

The fact that an inversion preserves sphere and planes may be easily reduced to the two-dimensional case (with circles and lines), a classical fact of Euclidean geometry. $\square$

2.1.8. The half-space model. We introduce another conformal model. The *half-space model* is the space

$$H^n = \big\{ (x_1, \ldots, x_n) \in \mathbb{R}^n \mid x_n > 0 \big\}.$$

It is obtained from the disc model D^n by an inversion in $\mathbb{R}^n$ with centre $C = (0, \ldots, 0, -1)$ and radius $\sqrt{2}$ as shown in Figure 2.6. The boundary ∂H^n is the horizontal hyperplane $\{x_n = 0\}$, to which we add the point ∞ at infinity to get a bijective correspondence between ∂H^n and ∂D^n through the inversion.

Proposition 2.1.22. *The half-space H^n is a conformal model for $\mathbb{H}^n$. Its k-subspaces are the k-planes and k-spheres in $\mathbb{R}^n$ orthogonal to ∂H^n.*
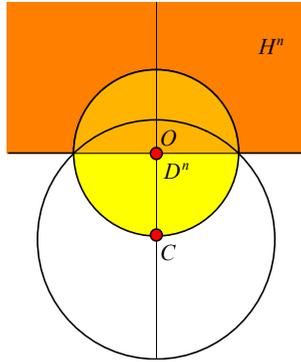
Figure 2.6. The inversion along the sphere with centre $C = (0, \ldots, 0, -1)$ and radius $\sqrt{2}$ transforms the Poincaré disc D^n into the half-space model H^n. Here $n = 2$.
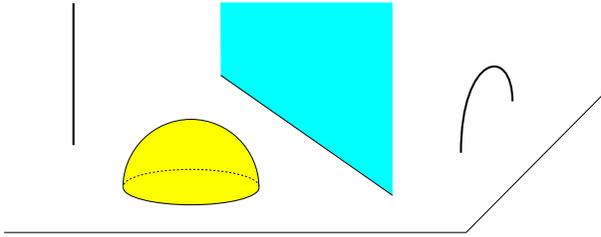
Figure 2.7. Lines and planes in the half-space model for $\mathbb{H}^3$.

Proof. The inversion is anticonformal and hence preserve angles, moreover it transforms k-spheres and k-planes in D^n orthogonal to ∂D^n into k-spheres and k-planes in H^n orthogonal to ∂H^n. □

Some lines and planes in H^3 are drawn in Figure 2.7. The metric tensor g on H^n has a particularly nice form.

Proposition 2.1.23. *The metric tensor on H^n is:*

$$g_x = \frac{1}{x_n^2} \cdot g^E$$

where g^E is the Euclidean metric tensor on $H^n \subset \mathbb{R}^n$.

Proof. The inversion $\varphi \colon D^n \to H^n$ is the function

$$\varphi(x_1, \ldots, x_n) = (0, \ldots, 0, -1) + 2\frac{(x_1, \ldots, x_{n-1}, x_n + 1)}{\|(x_1, \ldots, x_{n-1}, x_n + 1)\|^2}$$
$$= \frac{(2x_1, \ldots, 2x_{n-1}, 1 - \|x\|^2)}{\|x\|^2 + 2x_n + 1}.$$

As seen in the proof of Proposition 2.1.21, the inversion φ is anticonformal with scalar dilation

$$\frac{2}{\|(x_1, \ldots, x_{n-1}, x_n + 1)\|^2} = \frac{2}{\|x\|^2 + 2x_n + 1}.$$

The map φ hence transforms the metric tensor $\left(\frac{2}{1-\|x\|^2}\right)^2 \cdot g^E$ in $x \in D^n$ into the following metric tensor in $\varphi(x) \in H^n$:

$$\left(\frac{2}{1 - \|x\|^2}\right)^2 \cdot \left(\frac{\|x\|^2 + 2x_n + 1}{2}\right)^2 \cdot g^E = \frac{1}{\varphi_n(x)^2} \cdot g^E.$$

The proof is complete. □

2.1.9. Geodesics in the conformal models. The disc D^n and the half-space H^n are conformal models for $\mathbb{H}^n$. In both models the hyperbolic metric differs from the Euclidean one only by a multiplication with some function.

In the half-space H^n the lines are Euclidean vertical half-lines or half-circles orthogonal to ∂H^n as in Figure 2.7. Vertical geodesics have a particularly simple form.

Proposition 2.1.24. *The vertical geodesic in H^n passing through the point $(x_1, \ldots, x_{n-1}, 1)$ at time $t = 0$ pointing upward with unit speed is:*

$$\gamma(t) = (x_1, \ldots, x_{n-1}, e^t).$$

Proof. A tangent vector v at $(x_1, \ldots, x_n) \in H^n$ has norm $\frac{\|v\|^E}{x_n}$ where $\|v\|^E$ indicates the Euclidean norm. We get

$$\|\gamma'(t)\| = \|(0, \ldots, 0, e^t)\| = \frac{e^t}{e^t} = 1.$$

Therefore $\gamma(t)$ runs on a line at unit speed. □

We can easily deduce a parametrisation for the geodesics in D^n passing through the origin. Recall the hyperbolic tangent:

$$\tanh(t) = \frac{\sinh(t)}{\cosh(t)} = \frac{e^t - e^{-t}}{e^t + e^{-t}} = \frac{e^{2t} - 1}{e^{2t} + 1}.$$

Proposition 2.1.25. *The geodesic in D^n passing through the origin at time $t = 0$ and pointing towards $x \in S^{n-1}$ at unit speed is:*

$$\gamma(t) = \frac{e^t - 1}{e^t + 1} \cdot x = \left(\tanh \tfrac{t}{2}\right) \cdot x.$$

Proof. We can suppose $x = (0, \ldots, 0, 1)$ and obtain this parametrisation from that of the vertical line in H^n through inversion. □

We obtain in particular:

Corollary 2.1.26. *The exponential map* $\exp_0 \colon T_0 D^n \to D^n$ *at the origin* $0 \in D^n$ *is the diffeomorphism:*

$$\exp_0(x) = \frac{e^{\|x\|} - 1}{e^{\|x\|} + 1} \cdot \frac{x}{\|x\|} = \left(\tanh \tfrac{\|x\|}{2} \right) \cdot \frac{x}{\|x\|}.$$

Since the isometries of $\mathbb{H}^n$ act transitively on points, we deduce that the exponential map at any $p \in \mathbb{H}^n$ is a diffeomorphism. As a consequence, the injectivity radius of $\mathbb{H}^n$ is ∞, as in the Euclidean $\mathbb{R}^n$.

2.1.10. Isometries of the conformal models. In the half-space model it is easy to identify some isometries:

Proposition 2.1.27. *The following are isometries of H^n:*
- *horizontal translations $x \mapsto x + b$ with $b = (b_1, \ldots, b_{n-1}, 0)$,*
- *dilations $x \mapsto \lambda x$ with $\lambda > 0$,*
- *inversions with respect to spheres orthogonal to ∂H^n.*

Proof. Horizontal translations obviously preserve the metric tensor $g = \frac{1}{x_n^2} \cdot g^E$. We indicate by $\|\cdot\|$ and $\|\cdot\|^E$ the hyperbolic and Euclidean norm of tangent vectors. On a dilation $\varphi(x) = \lambda x$ we get

$$\|d\varphi_x(v)\| = \frac{\|d\varphi_x(v)\|^E}{\varphi(x)_n} = \frac{\lambda \|v\|^E}{\lambda x_n} = \frac{\|v\|^E}{x_n} = \|v\|.$$

Concerning inversions, up to conjugating by translations and dilations it suffices to consider the map $\varphi(x) = \frac{x}{\|x\|^2}$. We have already seen that $d\varphi_x$ is $\frac{1}{\|x\|^2}$ times a linear reflection. Therefore

$$\|d\varphi_x(v)\| = \frac{\|d\varphi_x(v)\|^E}{\varphi(x)_n} = \frac{\|v\|^E/\|x\|^2}{x_n/\|x\|^2} = \frac{\|v\|^E}{x_n} = \|v\|.$$

This completes the proof. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

In the disc model we can easily write the isometries that fix the origin:

Proposition 2.1.28. *The group $O(n)$ acts isometrically on D^n.*

Proof. The metric tensor on D^n has a spherical symmetry. $\qquad\qquad\square$

It is harder to write the isometries that fix another point of D^n or H^n.

Proposition 2.1.29. *On the conformal models D^n and H^n, the isometry group is generated by inversions along spheres and reflections along Euclidean planes orthogonal to the boundary.*

Proof. We know from Proposition 2.1.27 that these maps are isometries of H^n. One such isometry fixes a hyperplane S, hence it is the hyperbolic reflection r_S (which is the unique non-trivial isometry fixing S). Hyperbolic reflections generate the isometry group by Proposition 2.1.11.

In D^n the proof is the same, we leave as an exercise to prove that a sphere inversion preserves the metric tensor and is hence an isometry. $\quad\square$
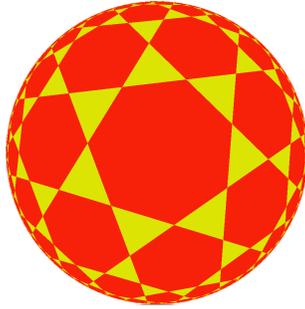
Figure 2.8. A tessellation of $\mathbb{H}^2$ into regular triangles and heptagons in the Klein model.

2.1.11. Balls in the conformal models. How does a metric ball in the hyperbolic space look like? The answer in the conformal models is surprisingly simple.

Proposition 2.1.30. *In the conformal models balls are Euclidean balls.*

Proof. In the disc model $B(0, r)$ is a ball centred at 0 by symmetry. The ball $B(x_0, r)$ at another point x_0 is obtained from $B(0, r)$ by composing inversions, which send spheres to spheres and hence balls to balls. The inversion $D^n \to H^n$ also sends balls to balls. □

The centre of the ball is not its Euclidean centre in general!

2.1.12. The Klein model. There is a fourth model for the hyperbolic space that is some kind of intermediate version between the hyperboloid and the Poincaré disc model.

The *Klein* or *projective model* for $\mathbb{H}^n$ is obtained by embedding the hyperboloid I^n inside $\mathbb{RP}^n$ via the projection $\mathbb{R}^{n+1} \setminus \{0\} \to \mathbb{RP}^n$. The image of this embedding is an open disc $K^n \subset \mathbb{RP}^n$ bounded by the quadric $x_1^2 + \ldots + x_n^2 - x_{n+1}^2 = 0$. We equip K^n with the metric tensor transported from I^n, so that K^n is indeed a model for $\mathbb{H}^n$.

When read in the chart $x_{n+1} = 1$, the Klein model K^n becomes the open disc $x_1^2 + \ldots + x_n^2 < 1$. It is like the Poincaré disc D^n, but with a different metric tensor!

The Klein model K^n is not conformal and its metric tensor is a bit more complicated than that of D^n or H^n. On the other hand, the subspaces in K^n are easier to identify: by definition, these are just the projective subspaces of $\mathbb{RP}^n$ intersected with the open disc K^n. Therefore lines are straight lines, but the angles and distances are distorted. Note also that the isometries of K^n are projective transformations. The distance function is particularly nice:

Exercise 2.1.31. The distance of two distinct points p, q in K^n is

$$d(p, q) = \frac{1}{2} |\log \beta(p, q, r, s)|$$

where r, s are the intersections of the projective line l containing p and q with ∂K^n, and $\beta(p, q, r, s)$ is the cross-ratio of the four points.

2.2. Compactification and isometries of hyperbolic space

In this section we compactify the hyperbolic space $\mathbb{H}^n$ by adding some "points at infinity". The compactification will then be used to classify the isometries of $\mathbb{H}^n$ into three types. We also study the mutual position of two subspaces and define some peculiar hyperbolic objects: the horospheres.

2.2.1. Points at infinity. Let a *geodesic half-line* in $\mathbb{H}^n$ be a geodesic $\gamma \colon [0, +\infty) \to \mathbb{H}^n$ with constant unit speed.

Definition 2.2.1. The set $\partial \mathbb{H}^n$ of the *points at infinity* in $\mathbb{H}^n$ is the set of all geodesic half-lines, taken up to the following equivalence relation:

$$\gamma_1 \sim \gamma_2 \iff \sup_{t \in [0, +\infty)} \left\{ d\big(\gamma_1(t), \gamma_2(t)\big) \right\} < +\infty.$$

We add its points at infinity to $\mathbb{H}^n$ by defining

$$\overline{\mathbb{H}^n} = \mathbb{H}^n \cup \partial \mathbb{H}^n.$$

Proposition 2.2.2. *On the disc model there is a natural 1-1 correspondence between ∂D^n and $\partial \mathbb{H}^n$ and hence between $\overline{D^n}$ and $\overline{\mathbb{H}^n}$.*

Proof. A geodesic half-line γ in D^n is a circle or line arc orthogonal to ∂D^n and hence the Euclidean limit $\lim_{t \to \infty} \gamma(t)$ is a point in ∂D^n. We now prove that two half-lines converge to the same point if and only if they lie in the same equivalence class.

Suppose that two half-geodesics γ_1, γ_2 converge to the same point $p \in \partial D^n$. Up to isometries and inversions, we can change the conformal model to H^n and take $p = \infty$. In this nicer setting γ_1 and γ_2 are vertical lines:

$$\gamma_1(t) = (x_1, \ldots, x_{n-1}, x_n e^t), \quad \gamma_2(t) = (y_1, \ldots, y_{n-1}, y_n e^t).$$

The geodesic

$$\gamma_3(t) = (y_1, \ldots, y_{n-1}, x_n e^t)$$

is equivalent to γ_2 since $d(\gamma_2(t), \gamma_3(t)) = |\ln \frac{y_n}{x_n}|$ for all t and is also equivalent to γ_1 because $d(\gamma_1(t), \gamma_3(t)) \to 0$ as shown in Figure 2.9.

Suppose that γ_1 and γ_2 converge to distinct points in ∂D^n. We can use the half-space model again and suppose that γ_1 is upwards vertical and γ_2 tends to some other point in $\{x_n = 0\}$. In that case we easily see that $d(\gamma_1(t), \gamma_2(t)) \to \infty$: for any $M > 0$ there is a $t_0 > 0$ such that $\gamma_1(t)$ and $\gamma_2(t)$ lie respectively in $\{x_n > M\}$ and $\{x_n < \frac{1}{M}\}$ for all $t > t_0$. Whatever curve connects these two open sets, it has length at least $\ln M^2$, hence $d(\gamma_1(t), \gamma_2(t)) > \ln M^2$ for all $t > t_0$. $\qquad \square$
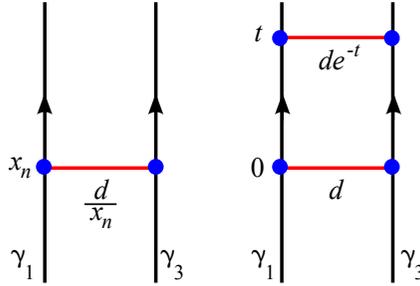
Figure 2.9. Two vertical lines γ_1 and γ_3 in the half-space model H^n at *Euclidean* distance d. The *hyperbolic* length of the horizontal segment between them at height x_n is $\frac{d}{x_n}$ and hence tends to zero as $x_n \to \infty$ (left). Using as a height parameter the more intrinsic hyperbolic arc-length, we see that the two vertical geodesics γ_1 and γ_3 approach at exponential rate, since $d(\gamma_1(t), \gamma_3(t)) \leqslant de^{-t}$ (right).

2.2.2. The compactification. We can give $\overline{\mathbb{H}^n}$ the topology of $\overline{D^n}$, and in that way we have *compactified* $\mathbb{H}^n$ by adding its points at infinity. The interior of $\overline{\mathbb{H}^n}$ is $\mathbb{H}^n$, and the points at infinity form a sphere $\partial \mathbb{H}^n$.

The topology on $\overline{\mathbb{H}^n}$ may also be defined intrinsically: for any $p \in \partial \mathbb{H}^n$ we define a system of open neighbourhoods of p in $\overline{\mathbb{H}^n}$ as follows. Let γ be a half-line with $[\gamma] = p$ and V be an open neighbourhood of the vector $\gamma'(0)$ in the unitary sphere in $T_{\gamma(0)}\mathbb{H}^n$. Pick $r > 0$ and define the following subset of $\overline{\mathbb{H}^n}$:

$$U(\gamma, V, r) = \big\{ \alpha(t) \mid \alpha(0) = \gamma(0),\ \alpha'(0) \in V,\ t > r \big\}$$
$$\bigcup \big\{ [\alpha] \mid \alpha(0) = \gamma(0),\ \alpha'(0) \in V \big\}$$

where α indicates a half-line in $\mathbb{H}^n$ and $[\alpha] \in \partial \mathbb{H}^n$ its class, see Figure 2.10. We define an open neighbourhoods system $\{U(\gamma, V, r)\}$ for p by letting γ, V, and r vary. The resulting topology on $\overline{\mathbb{H}^n}$ coincides with that of $\overline{D^n}$.

Note that $\mathbb{H}^n$ is a complete Riemannian manifold (and hence a complete metric space), while its compactification $\overline{\mathbb{H}^n}$ is only a topological space: a point in $\partial \mathbb{H}^n$ has infinite distance from any other point in $\overline{\mathbb{H}^n}$.

2.2.3. Klein model. In the Klein model $K^n \subset \mathbb{RP}^n$, the compactification is obtained by adding the quadric $\partial K^n = \{x_1^2 + \ldots + x_n^2 - x_{n+1}^2 = 0\}$, which is the image of the light cone in $\mathbb{R}^{n+1}$. Analogously, in the hyperboloid model I^n we may represent ∂I^n as the set of rays in the light cone.

A nice feature of the Klein model is that the points $x \in \mathbb{RP}^n$ that lie "beyond the infinity", that is outside $\overline{K^n}$, can also be given a natural geometric interpretation. One such $x \in \mathbb{RP}^n \setminus \overline{K^n}$ is a space-like ray in
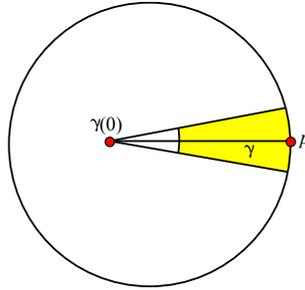
Figure 2.10. An open neighbourhood $U(\gamma, V, r)$ of $p \in \partial\mathbb{H}^n$ (in yellow). We use the Klein model here.
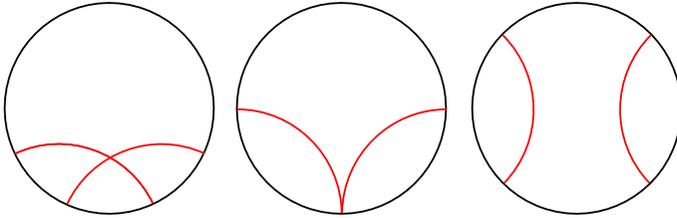
Figure 2.11. Two incident, asymptotic parallel, and ultra-parallel lines.

$\mathbb{R}^{n+1}$ and as such it defines a hyperplane $x^\perp$ of signature $(n-1, 1)$ in $\mathbb{R}^{n+1}$, which projects to a hyperbolic hyperplane in K^n. The points beyond the infinity are in 1-1 correspondence with the hyperbolic hyperplanes in K^n.

2.2.4. Incident, parallel, and ultraparallel subspaces. A k-subspace $S \subset \mathbb{H}^n$ has a topological closure $\overline{S}$ in the compactification $\overline{\mathbb{H}^n}$. The boundary $\partial S = \overline{S} \cap \partial\mathbb{H}^n$ of S is a $(k-1)$-sphere.

For instance, the boundary of a line l consists of two distinct points, the *endpoints* of l. The boundary of a plane is a circle. Two distinct points in $\partial\mathbb{H}^n$ are the endpoints of a unique line. A circle in the sphere ∂D^3 is the boundary of a unique plane in the disc model.

The usual *distance* $d(A, B)$ between two subsets A, B in a metric space is defined as

$$d(A, B) = \inf_{x \in A, y \in B} \{d(x, y)\}.$$

There are three types of configurations for two subspaces in $\mathbb{H}^n$, depicted in Figure 2.11.

Proposition 2.2.3. *Let S and S' be subspaces in $\mathbb{H}^n$ of arbitrary dimension. Precisely one of the following holds:*
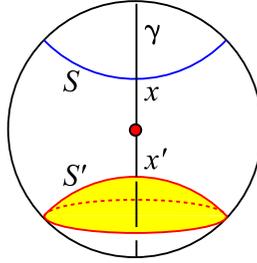
Figure 2.12. Two ultraparallel subspaces S, S' and a line γ orthogonal to both.

(1) $S \cap S' \neq \varnothing$,
(2) $S \cap S' = \varnothing$ and $\overline{S} \cap \overline{S'}$ is a point in $\partial \mathbb{H}^n$; moreover $d(S, S') = 0$ and there is no geodesic orthogonal to both S and S',
(3) $\overline{S} \cap \overline{S'} = \varnothing$; moreover $d = d(S, S') > 0$ and there is a unique geodesic γ orthogonal to both S, S': the segment of γ between S and S' is the unique arc connecting them of length d.

Proof. If $\overline{S} \cap \overline{S'}$ contains two points then it contains the line connecting them and hence $S \cap S' \neq \varnothing$.

In (2) we use the half-space model and send $\overline{S} \cap \overline{S'}$ at infinity. Then S and S' are Euclidean vertical subspaces and Figure 2.9 shows that $d(S, S') = 0$. Geodesics are vertical or half-circles and cannot be orthogonal to both S and S'.

In (3), let $x_i \in S$ and $x_i' \in S'$ be such that $d(x_i, x_i') \to d$. Since $\overline{\mathbb{H}^n}$ is compact, on a subsequence $x_i \to x \in \overline{S}$ and $x_i' \to x' \in \overline{S'}$. By hypothesis $x \neq x'$ and hence $x, x' \in \mathbb{H}^n$ since $d < \infty$. Therefore $d > 0$.

Let γ be the line passing through x and x'. The segment between x and x' has length $d(x, x') = d$. The line is orthogonal to S and S': if it had an angle smaller than $\frac{\pi}{2}$ with S' we could find another point $x'' \in S'$ near x' with $d(x, x'') < d$. We can draw S, S', γ as in Figure 2.12 by placing the origin between x and x': no other line can be orthogonal to both S and S'. $\square$

Two subspaces of type (1), (2) or (3) are called respectively *incident*, *asymptotically parallel*, and *ultra-parallel*.

2.2.5. The conformal sphere at infinity. The sphere at infinity $\partial \mathbb{H}^n$ has no metric structure, but it has instead a *conformal structure*, that is a Riemannian structure considered up to conformal transformations. Before defining it we note the following.

Proposition 2.2.4. *Every isometry $\varphi \colon \mathbb{H}^n \to \mathbb{H}^n$ extends to a unique homeomorphism $\varphi : \overline{\mathbb{H}^n} \to \overline{\mathbb{H}^n}$. An isometry φ is determined by its trace $\varphi|_{\partial \mathbb{H}^n}$ at the boundary.*

Proof. The extension of φ to $\partial \mathbb{H}^n$ is defined intrinsically: a boundary point is a class $[\gamma]$ of geodesic half-lines and we set $\varphi([\gamma]) = [\varphi(\gamma)]$.

To prove the second assertion we show that an isometry φ that fixes the points at infinity is the identity. The isometry φ fixes every line as a set (because it fixes its endpoints), and since every point is the intersection of two lines it fixes also every point. $\square$

We give $\partial \mathbb{H}^n$ the conformal structure of the sphere ∂D^n. The group $\mathrm{Isom}(\mathbb{H}^n)$ is generated by sphere inversions, which act conformally: although the metric tensor of ∂D^n is not preserved by this action, its conformal class is preserved and hence the conformal structure on $\partial \mathbb{H}^n$ is well-defined.

2.2.6. Elliptic, parabolic, and hyperbolic isometries. It is convenient to classify the isometries of $\mathbb{H}^n$ into three types.

Proposition 2.2.5. *Let φ be a non-trivial isometry of $\mathbb{H}^n$. Precisely one of the following holds:*

(1) *φ has at least one fixed point in $\mathbb{H}^n$,*
(2) *φ has no fixed points in $\mathbb{H}^n$ and has exactly one in $\partial \mathbb{H}^n$,*
(3) *φ has no fixed points in $\mathbb{H}^n$ and has exactly two in $\partial \mathbb{H}^n$.*

Proof. The extension $\varphi : \overline{\mathbb{H}^n} \to \overline{\mathbb{H}^n}$ is continuous and has a fixed point by Brouwer's Theorem. We only need to prove that if φ has three fixed points P_1, P_2, P_3 at the boundary then it has some fixed point in the interior. The isometry φ fixes the line r with endpoints P_1 and P_2. There is only one line s with endpoint P_3 and orthogonal to r (exercise): the isometry φ must also fix s and hence fixes the point $r \cap s$. $\square$

Isometries of type (1), (2), and (3) are called respectively *elliptic*, *parabolic*, and *hyperbolic*. A hyperbolic isometry fixes two points $p, q \in \partial \mathbb{H}^n$ and hence preserves the unique line l with endpoints p and q. The line l is the *axis* of the hyperbolic isometry, which acts on l as a translation.

2.2.7. Horospheres. Parabolic transformations are related to some objects in $\mathbb{H}^n$ called horospheres.

Definition 2.2.6. Let p be a point in $\partial \mathbb{H}^n$. A *horosphere* centred in p is a connected complete hypersurface orthogonal to all the lines exiting from p.

Horospheres may be easily visualised in the half-space model H^n by sending p at infinity. The lines exiting from p are the Euclidean vertical lines and the horospheres centred at p are precisely the horizontal hyperplanes $\{x_n = k\}$ with $k > 0$.
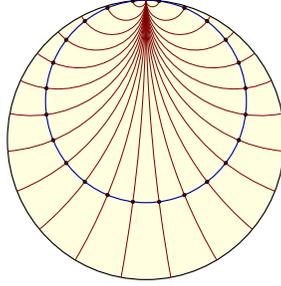
Figure 2.13. A horocycle in $\mathbb{H}^2$ centred in $p \in \partial\mathbb{H}^2$ is a circle tangent to p. It is orthogonal to all the lines exiting from p.

Remark 2.2.7. Since the metric tensor $g = \frac{1}{x_n^2} g^E$ is constant on each hyperplane $\{x_n = k\}$, each horosphere is *isometric* to the Euclidean $\mathbb{R}^n$.

The horospheres centred at $p \neq \infty$ in ∂H^n or at any point $p \in \partial D^n$ are precisely the Euclidean spheres tangent in p to the sphere at infinity. The horospheres in $\mathbb{H}^2$ are circles and are called *horocycles*, see Figure 2.13. The portion of $\mathbb{H}^n$ delimited by a horosphere is called a *horoball*.

Let us go back to the isometries of $\mathbb{H}^n$. We sometimes write a point in the half-space H^n as a pair (x, t) with $x \in \mathbb{R}^{n-1}$ and $t > 0$. Isometries with nice fixed points have nice expressions in the conformal models.

Proposition 2.2.8. *Let φ be an isometry of $\mathbb{H}^n$:*

(1) *if φ is elliptic with fixed point $0 \in D^n$ then*

$$\varphi(x) = Ax$$

for some matrix $A \in O(n)$;

(2) *if φ is parabolic with fixed point ∞ in H^n then*

$$\varphi(x, t) = (Ax + b, t)$$

for some matrix $A \in O(n-1)$ and some vector b;

(3) *if φ is hyperbolic with fixed points 0 and ∞ in H^n then*

$$\varphi(x, t) = \lambda(Ax, t)$$

for some matrix $A \in O(n-1)$ and some positive scalar $\lambda \neq 1$.

Proof. Point (1) is obvious. In (2) the isometry φ fixes ∞ and hence permutes the horospheres centred at ∞: we first prove that this permutation is trivial. The map φ sends a horosphere O_0 at height $t = t_0$ to a horosphere O_1 at some height $t = t_1$. If $t_1 \neq t_0$, up to changing φ with its inverse we may suppose that $t_1 < t_0$.

We know that the map $\psi \colon O_1 \to O_0$ sending (x, t_1) to (x, t_0) is a contraction: hence $\varphi \circ \psi \colon O_1 \to O_1$ is a contraction and thus has a fixed point (x, t_1). Therefore $\varphi(x, t_0) = (x, t_1)$. Since $\varphi(\infty) = \infty$, the vertical geodesic passing through (x, t_0) and (x, t_1) is preserved by φ, and hence we have found another fixed point $(x, 0) \in \partial \mathbb{H}^n$, a contradiction.

We now know that φ preserves the horosphere O at height t, for all t. The metric tensor on O is Euclidean (rescaled by $\frac{1}{t^2}$), hence φ acts on O like an isometry $x \mapsto Ax + b$. Since φ sends vertical geodesics to vertical geodesics, it acts with the same formula on each horosphere and we are done.

Concerning (3), the axis l of φ is the vertical line with endpoints $0 = (0, 0)$ and ∞, and φ acts on l by translations: hence it sends $(0, 1)$ to some $(0, \lambda)$. The differential $d\varphi$ at $(0, 1)$ is necessarily $\left(\begin{smallmatrix} A & 0 \\ 0 & \lambda \end{smallmatrix}\right)$ for some $A \in O(n-1)$ and hence φ is globally as stated. The case $\lambda = 1$ is excluded because $(0, 1)$ would be a fixed point in $\mathbb{H}^n$. $\square$

The *minimum displacement* $d = d(\varphi)$ of an isometry φ of $\mathbb{H}^n$ is

$$d(\varphi) = \inf_{x \in \mathbb{H}^n} d\big(x, \varphi(x)\big).$$

A point x realises the minimum displacement if $d(x, \varphi(x)) = d(\varphi)$.

Corollary 2.2.9. *The following hold:*
 (1) *an elliptic transformation φ has $d = 0$ realised on its fixed points;*
 (2) *a parabolic transformation φ with fixed point $p \in \partial \mathbb{H}^n$ has $d = 0$ realised nowhere and fixes every horosphere centred in p;*
 (3) *a hyperbolic transformation φ with fixed points $p, q \in \partial \mathbb{H}^n$ has $d > 0$ realised on its axis.*

Proof. Point (1) is obvious. Point (2) was already noticed while proving Proposition 2.2.8. Concerning (3), let l be the axis of the hyperbolic transformation φ. The hyperplane orthogonal to l at a point $x \in l$ is sent to the hyperplane orthogonal to l in $\varphi(x)$. The two hyerplanes are ultraparallel and by Proposition 2.2.3 their minimum distance is realised at the points x and $\varphi(x)$. Hence the points on l realise the minimum displacement for φ. $\square$

2.3. Isometry groups in dimensions two and three

With the hyperboloid model the isometry group $\mathrm{Isom}(\mathbb{H}^n)$ is the matrix group $O^+(n, 1)$. We now see that in dimensions $n = 2$ and 3 the group $\mathrm{Isom}^+(\mathbb{H}^n)$ is also isomorphic to some familiar groups of 2×2 matrices.

We start with a concise tour on Möbius transformations.

2.3.1. Möbius transformations. Let the *Riemann sphere* be $S = \mathbb{C} \cup \{\infty\}$, homeomorphic to S^2. Consider the group

$$\mathbb{PSL}_2(\mathbb{C}) = \mathrm{SL}_2(\mathbb{C})/_{\pm I} = \mathrm{GL}_2(\mathbb{C})/_{\{\lambda I\}} = \mathbb{PGL}_2(\mathbb{C})$$

of all 2×2 invertible complex matrices considered up to scalar multiplication. The group $\mathbb{PSL}_2(\mathbb{C})$ acts on S as follows: a matrix $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix} \right) \in \mathbb{PSL}_2(\mathbb{C})$ determines the *Möbius transformation*

$$z \mapsto \frac{az + b}{cz + d}$$

which is an orientation-preserving self-diffeomorphism of S, and in fact also a biolomorphism.

Exercise 2.3.1. Möbius transformations act freely and transitively on triples of distinct points in S.

A matrix $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix} \right) \in \mathbb{PSL}_2(\mathbb{C})$ also determines a *Möbius anti-transformation*

$$z \mapsto \frac{a\bar{z} + b}{c\bar{z} + d}$$

which is an orientation-reversing self-diffeomorphism of S. The composition of two anti-transformations is a Möbius transformation. Transformations and anti-transformations together form a group $\mathrm{Conf}(S)$ which contains the Möbius transformations as an index-two subgroup.

Proposition 2.3.2. *Circle inversions and line reflections are both Möbius anti-transformations and generate* $\mathrm{Conf}(S)$.

Proof. By conjugating with translations $z \mapsto z + b$ and complex dilations $z \mapsto az$ every circle inversion transforms into the inversion along the unit circle $z \mapsto \frac{1}{\bar{z}}$, and every line reflection transforms into $z \mapsto \bar{z}$.

By composing line reflections we get all translations and rotations, and by composing circle inversions we get all dilations. With these operations and the inversion $z \mapsto \frac{1}{\bar{z}}$ one can easily act transitively on triples of points. They generate $\mathrm{Conf}(S)$ by Exercise 2.3.1. $\qquad \square$

2.3.2. Möbius transformations of H^2. We consider the half-plane $H^2 \subset \mathbb{C}$ as $H^2 = \{z \mid \Im z > 0\}$ and denote by $\mathrm{Conf}(H^2)$ the subgroup of $\mathrm{Conf}(S)$ consisting of all maps that preserve H^2.

Remark 2.3.3. By standard results in complex analysis, the two groups $\mathrm{Conf}(S)$ and $\mathrm{Conf}(H^2)$ contain precisely all the conformal diffeomorphisms of S and H^2, whence their names. We will not use this fact here.

Proposition 2.3.4. *The maps in* $\mathrm{Conf}(H^2)$ *are of the form*

$$z \mapsto \frac{az + b}{cz + d} \qquad \text{and} \qquad z \mapsto \frac{a\bar{z} + b}{c\bar{z} + d}$$

with $a, b, c, d \in \mathbb{R}$ *and having* $ad - bc$ *equal to* 1 *and* -1, *respectively.*

Proof. The transformations listed have real coefficients and hence preserve the line $\mathbb{R} \cup \infty$ and permute the two half-planes in $\mathbb{C} \setminus \mathbb{R}$. The sign condition on $ad - bc = \pm 1$ ensures precisely that i is sent to some point in H^2 and hence H^2 is preserved.

On the other hand, a transformation that preserves H^2 must preserve $\partial H^2 = \mathbb{R} \cup \infty$ and it is easy to see that since the images of $0, 1, \infty$ are real all the coefficients a, b, c, d can be taken in $\mathbb{R}$. $\qquad\square$

The Möbius transformations in $\mathrm{Conf}(H^2)$ form a subgroup of index two which is naturally isomorphic to

$$\mathbb{P}\mathrm{SL}_2(\mathbb{R}) = \mathrm{SL}_2(\mathbb{R})/_{\pm I}.$$

An ordered triple of distinct points in $\mathbb{R} \cup \infty$ is *positive* if they are oriented counterclockwise, like $0, 1, \infty$.

Exercise 2.3.5. The group $\mathbb{P}\mathrm{SL}_2(\mathbb{R})$ acts freely and transitively on positive triples of points in $\mathbb{R} \cup \infty$.

Let $C \subset \mathbb{C}$ be a circle or line orthogonal to $\mathbb{R}$. The inversion or reflection along C preserves H^2 and is hence an element of $\mathrm{Conf}(H^2)$.

Proposition 2.3.6. *Inversions along circles and reflections along lines orthogonal to $\mathbb{R}$ generate $\mathrm{Conf}(H^2)$.*

Proof. Composing reflections we obtain all horizontal translations $z \mapsto z + b$ with $b \in \mathbb{R}$, composing inversions we obtain all dilations $z \mapsto \lambda z$ with $\lambda \in \mathbb{R}^*$. These maps together with the inversion $z \mapsto \frac{1}{\bar{z}}$ act transitively on positive triples of points in $\mathbb{R} \cup \infty$. $\qquad\square$

Exercise 2.3.7. The inversion sending H^2 to D^2 is

$$z \mapsto \frac{\bar{z} + i}{i\bar{z} + 1}.$$

2.3.3. Isometries of H^2. After this short detour on Möbius transformations, we turn back to our hyperbolic spaces. We can characterise the isometry group of H^2.

Proposition 2.3.8. *We have $\mathrm{Isom}(H^2) = \mathrm{Conf}(H^2)$.*

Proof. Both groups are generated by inversions along circles and reflections along lines orthogonal to $\partial H^2 = \mathbb{R}$ by Propositions 2.1.29 and 2.3.6. $\qquad\square$

In particular we have

$$\mathrm{Isom}^+(H^2) = \mathbb{P}\mathrm{SL}_2(\mathbb{R}).$$

We will henceforth identify these two groups. The trace of an element in $\mathbb{P}\mathrm{SL}_2(\mathbb{R})$ is well-defined up to sign and carries some relevant information:

Proposition 2.3.9. *A non-trivial isometry $A \in \mathbb{P}\mathrm{SL}_2(\mathbb{R})$ is elliptic, parabolic, hyperbolic $\iff$ respectively $|\mathrm{tr}A| < 2$, $|\mathrm{tr}A| = 2$, $|\mathrm{tr}A| > 2$.*

Proof. Take $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ with $\det A = ad - bc = 1$. The Möbius transformation $z \mapsto \frac{az+b}{cz+d}$ has a fixed point $z \in \mathbb{C}$ if and only if

$$\frac{az+b}{cz+d} = z \Longleftrightarrow cz^2 + (d - a)z - b = 0.$$

We find

$$\Delta = (d - a)^2 + 4bc = (d + a)^2 - 4 = \mathrm{tr}^2 A - 4.$$

There is a fixed point in H^2 if and only if $\Delta < 0$; if $\Delta > 0$ we find two fixed points in $\mathbb{R} \cup \{\infty\}$ and if $\Delta = 0$ only one. $\square$

2.3.4. Isometries of H^3. We have proved that $\mathrm{Isom}^+(H^2) = \mathbb{PSL}_2(\mathbb{R})$. Quite surprisingly, the isometry group $\mathrm{Isom}^+(H^3)$ is also isomorphic to a group of 2×2 matrices! To prove this, we make the following identifications:

$$\mathbb{R}^3 = \mathbb{C} \times \mathbb{R} = \{(z, t) \mid z \in \mathbb{C}, t \in \mathbb{R}\}$$

hence $H^3 = \{(z, t) \mid t > 0\}$. We also write $\mathbb{C}$ for $\mathbb{C} \times \{0\}$. The boundary trace of an isometry of H^3 is a homeomorphism of the Riemann sphere

$$\partial H^3 = \mathbb{C} \cup \{\infty\} = S.$$

Proposition 2.3.10. *The boundary trace induces an identification*

$$\mathrm{Isom}(H^3) = \mathrm{Conf}(S).$$

Proof. The group $\mathrm{Isom}(H^3)$ is generated by inversions along spheres and reflections along planes orthogonal to ∂H^3. Their traces are inversions along circles and reflections along lines in S. These generate $\mathrm{Conf}(S)$. $\square$

In particular we have

$$\mathrm{Isom}^+(H^3) = \mathbb{PSL}_2(\mathbb{C}).$$

We will also henceforth identify these two groups. As above, the trace of an element in $\mathbb{PSL}_2(\mathbb{C})$ is well-defined up to sign and carries some information:

Proposition 2.3.11. *A non-trivial isometry $A \in \mathbb{PSL}_2(\mathbb{C})$ is elliptic, parabolic, hyperbolic if and only if respectively $\mathrm{tr} A \in (-2, 2)$, $\mathrm{tr} A = \pm 2$, $\mathrm{tr} A \in \mathbb{C} \setminus [-2, 2]$.*

Proof. Every non-trivial matrix $A \in \mathrm{SL}_2(\mathbb{C})$ is conjugate to one of:

$$\pm \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{pmatrix}$$

for some $\lambda \in \mathbb{C}^*$, and these represent the following isometries:

$$(z, t) \longmapsto (z + 1, t), \quad (z, t) \longmapsto (\lambda^2 z, |\lambda|^2 t).$$

In the first case $\mathrm{tr} A = \pm 2$ and A is parabolic with fixed point ∞, in the second case A has a fixed point in H^3 if and only if $|\lambda| = 1$, *i.e.* $\mathrm{tr} A = \lambda + \lambda^{-1} \in (-2, 2)$, the fixed point being $(0, 1)$. If $|\lambda| \neq 1$ there are two fixed points 0 and ∞ at infinity and hence A is hyperbolic. $\square$

Summing up, we have

$$\mathrm{Isom}^+(H^2) = \mathbb{PSL}_2(\mathbb{R}), \qquad \mathrm{Isom}^+(H^3) = \mathbb{PSL}_2(\mathbb{C}).$$

The group $\mathbb{PSL}_2(\mathbb{R})$ acts directly on H^2, while $\mathbb{PSL}_2(\mathbb{C})$ acts on the boundary sphere of H^3.

2.4. Geometry of hyperbolic space

We study the geometry of $\mathbb{H}^n$. We prove that $\mathbb{H}^n$ has constant sectional curvature -1, that the distance function along lines is convex, we define convex combinations and barycenters, and study parallel transport along lines. Finally, we prove that $\mathrm{Isom}(\mathbb{H}^n)$ is a unimodular Lie group.

2.4.1. Area and curvature. We can verify that $\mathbb{H}^n$ has constant sectional curvature -1. It should be no surprise that $\mathbb{H}^n$ has constant curvature, since it has many symmetries (*i.e.* isometries). To calculate its sectional curvature we compute the area of a disc.

Proposition 2.4.1. *The disc of radius r in $\mathbb{H}^2$ has area*

$$A(r) = \pi \left(e^{\frac{r}{2}} - e^{-\frac{r}{2}} \right)^2 = 4\pi \sinh^2 \tfrac{r}{2} = 2\pi(\cosh r - 1).$$

Proof. Recall that the volume form is

$$\omega = \sqrt{\det g} \cdot dx_1 \cdots dx_n.$$

Let $D(r)$ be a disc in $\mathbb{H}^2$ of radius r. If we centre it in 0 in the disc model, its Euclidean radius is $\tanh \frac{r}{2}$ by Corollary 2.1.26 and we get

$$A(r) = \int_{D(r)} \sqrt{\det g} \cdot dxdy = \int_{D(r)} \left(\frac{2}{1 - x^2 - y^2} \right)^2 dxdy$$

$$= \int_0^{2\pi} \int_0^{\tanh \frac{r}{2}} \left(\frac{2}{1 - \rho^2} \right)^2 \rho \cdot d\rho d\theta = 2\pi \left[\frac{2}{1 - \rho^2} \right]_0^{\tanh \frac{r}{2}}$$

$$= 4\pi \left(\frac{1}{1 - \tanh^2 \frac{r}{2}} - 1 \right) = 4\pi \sinh^2 \tfrac{r}{2}.$$

The proof is complete. $\qquad\qquad\square$

Corollary 2.4.2. *The hyperbolic space $\mathbb{H}^n$ has sectional curvature -1.*

Proof. Pick $p \in \mathbb{H}^n$ and $W \subset T_p$ a 2-dimensional subspace. The image $\exp_p(W)$ is the hyperbolic plane tangent to W in p. On a hyperbolic plane

$$A(r) = 2\pi(\cosh r - 1) = 2\pi \left(\frac{r^2}{2!} + \frac{r^4}{4!} + o(r^4) \right) = \pi r^2 + \frac{\pi r^4}{12} + o(r^4)$$

and hence $K = -1$ following the area formula in Section 1.2.8. $\qquad\square$
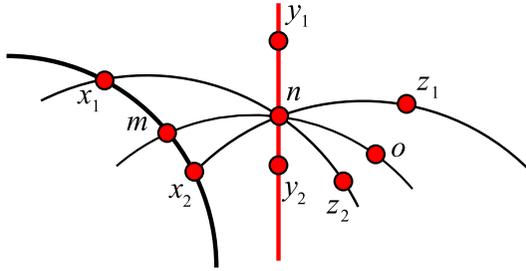
Figure 2.14. Distance between points in disjoint lines is a strictly convex function in hyperbolic space.

2.4.2. Convexity of the distance function. We recall that a function $f\colon \mathbb{R}^n \to \mathbb{R}$ is *strictly convex* if

$$f(tv + (1-t)w) < tf(v) + (1-t)f(w)$$

for any pair $v, w \in \mathbb{R}^n$ of distinct points and any $t \in (0,1)$.

Exercise 2.4.3. A positive strictly convex function is continuous and admits a minimum if and only if it is proper.

We now prove that the distance function is strictly convex on disjoint lines of $\mathbb{H}^n$. Given two lines $l, l' \subset \mathbb{H}^n$, we fix an isometry of each line with $\mathbb{R}$ and we get an identification of $l \times l'$ with the Euclidean plane $\mathbb{R} \times \mathbb{R}$.

Proposition 2.4.4. *Let* $l, l' \subset \mathbb{H}^n$ *be two disjoint lines. The map*

$$l \times l' \longrightarrow \mathbb{R}_{\geqslant 0}$$
$$(x, y) \longmapsto d(x, y)$$

is strictly convex; it is proper if and only if the lines are ultraparallel.

Proof. With our identifications we have $(x, y) \in \mathbb{R} \times \mathbb{R}$. The function d is clearly continuous, hence to prove its convexity it suffices to show that

$$d\left(\frac{x_1 + x_2}{2}, \frac{y_1 + y_2}{2}\right) < \frac{d(x_1, y_1) + d(x_2, y_2)}{2}$$

for any pair of distinct points $(x_1, y_1), (x_2, y_2) \in l \times l'$. Suppose $x_1 \neq x_2$ and denote by m and n the midpoints $\frac{x_1 + x_2}{2}$ and $\frac{y_1 + y_2}{2}$ as in Figure 2.14.

Let σ_p be the reflection at the point $p \in \mathbb{H}^n$. The isometry $\tau = \sigma_n \circ \sigma_m$ translates the line r containing the segment mn by the quantity $2d(m, n)$: hence it is a hyperbolic transformation with axis r. We draw the points $o = \tau(m)$ and $z_i = \tau(x_i)$ in the figure and note that $z_1 = \sigma_n(x_2)$, hence $d(x_2, y_2) = d(z_1, y_1)$. The triangular inequality implies that

$$d(x_1, z_1) \leqslant d(x_1, y_1) + d(y_1, z_1) = d(x_1, y_1) + d(x_2, y_2).$$

A hyperbolic transformation has minimum displacement on its axis r and $x_1 \neq m$ is not in r, hence

$$2d(m, n) = d(m, o) = d(m, \tau(m)) < d(x_1, \tau(x_1)) = d(x_1, z_1).$$

Finally we get $2d(m, n) < d(x_1, y_1) + d(x_2, y_2)$ and hence d is convex.

The function d is proper, that is it has minimum, if and only if the two lines are ultraparallel by Proposition 2.2.3. □

Exercise 2.4.5. The distance function on parallel lines in $\mathbb{R}^n$ is not strictly convex (it is only convex).

2.4.3. Convex combinations. Let $p_1, \ldots, p_k$ be k points in $\mathbb{H}^n, \mathbb{R}^n$, or S^n and $t_1, \ldots, t_k$ be non-negative numbers with $t_1 + \ldots + t_k = 1$. The *convex combination*

$$p = t_1 p_1 + \ldots + t_k p_k$$

is another point in the space defined as follows:

$$\text{in } \mathbb{R}^n: \quad p = t_1 p_1 + \ldots + t_k p_k$$

$$\text{in } I^n, S^n: \quad p = \frac{t_1 p_1 + \ldots + t_k p_k}{\| t_1 p_1 + \ldots + t_k p_k \|}$$

where $\|v\| = \sqrt{-\langle v, v \rangle}$ on I^n. Using convex combination we may define the *barycenter* of the points as $\frac{1}{k} p_1 + \ldots + \frac{1}{k} p_k$. The barycenter may in turn be used to prove the following.

Proposition 2.4.6. *Let $\varphi \colon \mathbb{H}^n \to \mathbb{H}^n$ be a non-trivial isometry and $k \geqslant 2$. Then:*

- *if φ is elliptic then φ^k is elliptic or trivial;*
- *if φ is parabolic then φ^k is parabolic;*
- *if φ is hyperbolic then φ^k is hyperbolic.*

Proof. If $\varphi \colon \mathbb{H}^n \to \mathbb{H}^n$ is an isometry with no fixed points, then φ^k also is: if $\varphi^k(x) = x$ then φ fixes the finite set $\{x, \varphi(x), \ldots, \varphi^{k-1}(x)\}$ and hence also its barycenter.

If φ is parabolic then it fixes the horospheres centred at some point $p \in \partial \mathbb{H}^n$ and also φ^k does, hence it is still parabolic (it cannot be hyperbolic). If φ is hyperbolic it has two fixed points at infinity, and φ^k too. □

Proposition 2.4.7. *Every finite subgroup $\Gamma < \mathrm{Isom}(\mathbb{H}^n)$ fixes a $p \in \mathbb{H}^n$.*

Proof. The barycenter of any orbit is fixed by Γ. □

2.4.4. Parallel transport. On Riemannian manifolds, the parallel transport is a way to slide frames along geodesics. On $\mathbb{H}^n$ we can do this simply as follows: for every geodesic γ, we put γ into vertical position in the half-space model H^n, and slide the frames vertically in the obvious way.

This construction furnishes in particular, for every pair $x, y \in \mathbb{H}^n$ of points, a canonical isometry between $T_x \mathbb{H}^n$ and $T_y \mathbb{H}^n$, obtained by sliding frames along the unique geodesic γ containing x and y.

This canonical identification is of course not transitive on a triple of non-collinear points x, y, z: the curvature of $\mathbb{H}^n$ is responsible for that.

2.4.5. Unimodularity. This is a consequence of Corollary 1.4.16.

Corollary 2.4.8. *The isometry group* $\mathrm{Isom}(\mathbb{H}^n)$ *is unimodular.*

Remark 2.4.9. A Haar measure for $\mathrm{Isom}(\mathbb{H}^n)$ may be constructed concretely as follows: fix a point $x \in \mathbb{H}^n$ and define the measure of a Borel set $S \subset \mathrm{Isom}(\mathbb{H}^n)$ as the measure of $S(x) = \cup_{\varphi \in S} \varphi(x) \subset \mathbb{H}^n$. This measure is obviously left-invariant, and is hence also right-invariant since $\mathrm{Isom}(\mathbb{H}^n)$ is unimodular. As a consequence, it does not depend on the choice of x.

2.4.6. References. The hyperbolic space is introduced in many books: two standard references are Benedetti − Petronio [**4**] and Ratcliffe [**49**], and most of the arguments presented here were borrowed from these two sources. The proof of Proposition 2.4.4 was taken from Farb − Margalit [**18**]. Thurston's notes contain some useful trigonometric formulae that we have omitted, see [**56**, Chapter 2].

CHAPTER 3

Hyperbolic manifolds

A hyperbolic manifold is a Riemannian manifold locally isometric to the hyperbolic space $\mathbb{H}^n$. Maybe the most striking aspect of geometric topology is that, despite this quite restrictive definition, there are plenty of hyperbolic manifolds around, especially in the dimensions $n = 2$ and 3. For that reason hyperbolic manifolds (and hence hyperbolic geometry) play a central role in the topology of surfaces and three-manifolds.

The study of complete hyperbolic manifolds is tightly connected to that of discrete subgroups in the Lie group $\text{Isom}(\mathbb{H}^n)$ and of polyhedra in $\mathbb{H}^n$, so it has both an algebraic and geometric flavour. We start this chapter by describing these connections; then we show some examples and discuss some important variations: non-complete hyperbolic manifolds, hyperbolic manifolds with geodesic boundary, cone manifolds, and orbifolds.

3.1. Discrete groups of isometries

We define hyperbolic manifolds and prove a crucial theorem, that says that every complete hyperbolic manifold is isometric to a quotient $\mathbb{H}^n/_\Gamma$ for some discrete group $\Gamma < \text{Isom}(\mathbb{H}^n)$ acting freely on $\mathbb{H}^n$.

3.1.1. Hyperbolic, flat, and elliptic manifolds. We introduce three important classes of Riemannian manifolds.

Definition 3.1.1. A *hyperbolic* (resp. *flat* or *elliptic*) *manifold* is a connected Riemannian n-manifold that may be covered by open sets isometric to open sets of $\mathbb{H}^n$ (resp. $\mathbb{R}^n$ o S^n).

A hyperbolic (resp. flat or elliptic) manifold has constant sectional curvature -1 (resp. 0 or $+1$). We show that the model $\mathbb{H}^n$ is indeed unique.

Theorem 3.1.2. *Every complete simply connected hyperbolic manifold M is isometric to $\mathbb{H}^n$.*

Proof. Pick a point $x \in M$ and choose an isometry $D \colon U \to V$ between an open ball U containing x and an open ball $V \subset \mathbb{H}^n$. We show that D extends (uniquely) to an isometry $D \colon M \to \mathbb{H}^n$.

For every $y \in M$, choose an arc $\alpha \colon [0, 1] \to M$ from x to y. By compactness there is a partition $0 = t_0 < t_1 < \ldots < t_k = 1$ and for

each $i = 0, \ldots, k - 1$ an isometry $D_i \colon U_i \to V_i$ from an open ball U_i in M containing $\alpha([t_i, t_{i+1}])$ to an open ball $V_i \subset \mathbb{H}^n$.

We may suppose that $U_0 \subset U$ and $D_0 = D|_{U_0}$. Inductively on i, we now modify D_i so that D_{i-1} and D_i coincide on the component C of $U_{i-1} \cap U_i$ containing $\alpha(t_i)$. To do so, note that

$$D_{i-1} \circ D_i^{-1} \colon D_i(C) \longrightarrow D_{i-1}(C)$$

is an isometry of open connected sets in $\mathbb{H}^n$ and hence extends to an isometry of $\mathbb{H}^n$. Then it makes sense to compose D_i with $D_{i-1} \circ D_i^{-1}$, so that the new maps D_{i-1} and D_i coincide on C. Finally, we define $D(y) = D_{k-1}(y)$.

The proof that $D(y)$ is well-defined is a standard argument. First, it is easy to check that different partitions $0 = t_0 < \ldots < t_k = 1$ do not vary $D(y)$, just by considering a common refinement. Then we consider another path β connecting x to y. Since M is simply-connected, there is a homotopy connecting α and β. The image of the homotopy is compact and is hence covered by finitely many open balls U_i isometric to open balls $V_i \subset \mathbb{H}^n$ via some maps D_i. By the Lebesgue number theorem, there is a $N > 0$ such that in the grid $[0, 1] \times [0, 1]$ of $\frac{1}{N} \times \frac{1}{N}$ squares, the image of every square is entirely contained in at least one U_i. We can now modify as above the isometries D_i inductively on the grid, starting from the bottom-left square, so that they all glue up and show that $D(y)$ does not depend on α or β.

The resulting map $D \colon M \to \mathbb{H}^n$ is a local isometry by construction. Since M is complete, the map D is a covering by Proposition 1.2.19. Since $\mathbb{H}^n$ is simply connected, the covering D is a homeomorphism and D is actually an isometry. $\square$

The isometry $D \colon M \to \mathbb{H}^n$ constructed in the proof is called a *developing map*. The same proof shows that every complete simply connected flat (or elliptic) n-manifold is isometric to $\mathbb{R}^n$ (or S^n).

3.1.2. Complete hyperbolic manifolds. We have determined the unique complete simply connected hyperbolic n-manifold, and we now look at complete hyperbolic manifolds with arbitrary fundamental group. We first note that if $\Gamma < \mathrm{Isom}(\mathbb{H}^n)$ is a group of isometries that acts freely and properly discontinuously on $\mathbb{H}^n$, the quotient manifold $\mathbb{H}^n/_\Gamma$ has a natural Riemannian structure that promotes the covering

$$\pi \colon \mathbb{H}^n \longrightarrow \mathbb{H}^n/_\Gamma$$

to a local isometry, see Proposition 1.5.9. The quotient $\mathbb{H}^n/_\Gamma$ is a complete hyperbolic manifold. We now show that every complete hyperbolic manifold is realised in this way:

Proposition 3.1.3. *Every complete hyperbolic n-manifold M is isometric to $\mathbb{H}^n/_\Gamma$ for some subgroup $\Gamma < \mathrm{Isom}(\mathbb{H}^n)$ acting freely and properly discontinuously.*

Proof. The universal cover of M inherits a Riemannian structure that is complete (by Proposition 1.2.19), hyperbolic, and simply connected: hence it is isometric to $\mathbb{H}^n$ by Theorem 3.1.2. The deck transformations Γ of the covering $\mathbb{H}^n \to M$ are necessarily locally isometries, hence isometries. We conclude that $M = \mathbb{H}^n/_\Gamma$ and Γ acts freely and properly discontinuously using Proposition 1.5.1. $\qquad\square$

Note that Γ is isomorphic to the fundamental group $\pi_1(M)$.

Remark 3.1.4. A group $\Gamma < \mathrm{Isom}(\mathbb{H}^n)$ acts freely if and only if it does not contain elliptic isometries: that is, every non-trivial isometry in Γ is either hyperbolic or parabolic.

Note also that Γ acts properly discontinuously if and only if it is discrete, see Proposition 1.5.8.

Remark 3.1.5. The same proofs show that every complete flat or spherical n-manifold is isometric to $\mathbb{R}^n/_\Gamma$ or $S^n/_\Gamma$ for some discrete group Γ of isometries acting freely on $\mathbb{R}^n$ or S^n.

Corollary 3.1.6. *There is a natural 1-1 correspondence*

$$\left\{ \begin{array}{c} \text{complete hyperbolic} \\ \text{manifolds } M \\ \text{up to isometry} \end{array} \right\} \longleftrightarrow \left\{ \begin{array}{c} \text{discrete subgroups } \Gamma < \mathrm{Isom}(\mathbb{H}^n) \\ \text{without elliptics} \\ \text{up to conjugation} \end{array} \right\}$$

Proof. When passing from the complete hyperbolic manifold M to the group Γ, the only choice we made is an isometry between the universal cover of M and $\mathbb{H}^n$. Different choices produce conjugate groups Γ. $\qquad\square$

3.1.3. Discrete groups. We investigate some basic properties of discrete groups Γ of isometries of $\mathbb{H}^n$.

Exercise 3.1.7. If $\Gamma < \mathrm{Isom}(\mathbb{H}^n)$ is discrete then it is countable.

Note that Γ is not necessarily finitely generated. We denote by $\Gamma_p < \Gamma$ the stabiliser of a point $p \in \mathbb{H}^n$.

Proposition 3.1.8. *Let $\Gamma < \mathrm{Isom}(\mathbb{H}^n)$ be discrete and $p \in \mathbb{H}^n$ a point. The stabiliser Γ_p is finite and the orbit $\Gamma(p) = \{g(p) \mid g \in \Gamma\}$ is discrete.*

Proof. Both are obvious consequence of the fact that Γ acts properly discontinuously. $\qquad\square$

Of course Γ acts freely on $\mathbb{H}^n$ if and only if Γ_p is trivial for all $p \in \mathbb{H}^n$. A set of subspaces in $\mathbb{H}^n$ is *locally finite* if every compact subset in $\mathbb{H}^n$ intersects only finitely many of them.

Proposition 3.1.9. *Let $\Gamma < \mathrm{Isom}(\mathbb{H}^n)$ be discrete. The points $p \in \mathbb{H}^n$ with trivial stabiliser Γ_p form an open dense set in $\mathbb{H}^n$.*

Proof. The fixed-points set $\mathrm{Fix}(g)$ of a non-trivial isometry g is a proper subspace of $\mathbb{H}^n$. The subspaces $\mathrm{Fix}(g)$ with $g \in \Gamma$ are locally finite: if infinitely many of them intersect a compact set they accumulate and Γ does not act properly discontinuously. The complement of a locally finite set of proper subspaces is open and dense. $\qquad\square$

Recall that a group *has no torsion* if every non-trivial element has infinite order.

Proposition 3.1.10. *A discrete group $\Gamma < \mathrm{Isom}(\mathbb{H}^n)$ acts freely on $\mathbb{H}^n$ if and only if it has no torsion.*

Proof. By Proposition 2.4.6 parabolic and hyperbolic elements have infinite order. On the other hand, an elliptic element $g \in \Gamma$ has finite order since Γ_p is finite for $p \in \mathrm{Fix}(g)$. $\qquad\square$

Corollary 3.1.11. *The fundamental group of a complete hyperbolic manifold has no torsion.*

It is now time to exhibit some examples. Recall that H^2 is the half-plane model, see Section 2.3.

Example 3.1.12. The *modular group*

$$\Gamma = \mathbb{PSL}_2(\mathbb{Z}) < \mathbb{PSL}_2(\mathbb{R}) = \mathrm{Isom}^+(H^2)$$

consists of all matrices in $\mathbb{PSL}_2(\mathbb{R})$ having integer entries and is clearly a discrete subgroup. It does not act freely on H^2, however: the matrix $\left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)$ represents the elliptic transformation $z \mapsto -\frac{1}{z}$ with fixed point i.

3.1.4. Coverings. We now make a simple but crucial observation: if $\Gamma < \mathrm{Isom}(\mathbb{H}^n)$ acts freely and properly discontinuously, then also every subgroup $\Gamma' < \Gamma$ does; we get a manifolds covering

$$\mathbb{H}^n/_{\Gamma'} \longrightarrow \mathbb{H}^n/_\Gamma$$

whose degree d is precisely the index of Γ' in Γ. Recall from Proposition 1.2.20 that we get

$$\mathrm{Vol}\big(\mathbb{H}^n/_{\Gamma'}\big) = d \cdot \mathrm{Vol}\big(\mathbb{H}^n/_\Gamma\big)$$

where some of the terms in the formula may be infinite. Moreover, every covering of a hyperbolic complete $M = \mathbb{H}^n/_\Gamma$ is constructed in this way: there is a nice bijective correspondence

$$\big\{\text{coverings of } M\big\} \longleftrightarrow \big\{\text{subgroups of } \Gamma\big\}.$$

This holds of course also for flat and spherical manifolds.

3.1.5. Congruence subgroups. We can now exhibit a family of two-dimensional hyperbolic manifolds.

Pick an integer $m \geqslant 2$. Let $\mathrm{SL}_2(\mathbb{Z}/_{m\mathbb{Z}})$ be the group of 2×2 matrices with coefficients in $\mathbb{Z}/_{m\mathbb{Z}}$ and determinant 1. We define the quotient

$$\mathbb{PSL}_2(\mathbb{Z}/_{m\mathbb{Z}}) = \mathrm{SL}_2(\mathbb{Z}/_{m\mathbb{Z}})/_{\{\pm I\}}.$$

The reduction modulo m homomorphism $\mathbb{Z} \to \mathbb{Z}/_{m\mathbb{Z}}$ induces the group homomorphisms $\mathrm{SL}_2(\mathbb{Z}) \to \mathrm{SL}_2(\mathbb{Z}/_{m\mathbb{Z}})$ and

$$\mathbb{PSL}_2(\mathbb{Z}) \longrightarrow \mathbb{PSL}_2(\mathbb{Z}/_{m\mathbb{Z}}).$$

The kernel of this homomorphism is the *principal congruence subgroup* $\Gamma(m)$ of $\mathbb{PSL}_2(\mathbb{Z})$. It is clearly discrete, since $\mathbb{PSL}_2(\mathbb{Z})$ is. It has finite index in $\mathbb{PSL}_2(\mathbb{Z})$ because $\mathbb{PSL}_2(\mathbb{Z}/_{m\mathbb{Z}})$ is finite.

Proposition 3.1.13. *If $m \geqslant 4$ the group $\Gamma(m)$ acts freely on H^2.*

Proof. An element $A \in \Gamma(m)$ is a matrix $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ congruent to $\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$ modulo m. In particular $a + d$ is congruent to 2 modulo m, and hence is not $-1, 0, 1$. Therefore A is never elliptic. $\qquad\square$

The quotient $\mathbb{H}^2/_{\Gamma(m)}$ is a hyperbolic surface. We will construct many hyperbolic surfaces in Section 6.2 via some more geometric methods.

3.1.6. Selberg's lemma. The discrete group $\Gamma = \mathbb{PSL}_2(\mathbb{Z})$ does not act freely on H^2, but its finite-index normal subgroup $\Gamma(m)$ does as soon as $m \geqslant 4$. Is this the instance of a more general principle? Yes, it is.

Proposition 3.1.14. *Every finitely generated discrete subgroup $\Gamma < \mathrm{Isom}(\mathbb{H}^n)$ has a finite-index normal subgroup $\Gamma' \lhd \Gamma$ that acts freely on $\mathbb{H}^n$.*

Proof. The group $\mathrm{Isom}(\mathbb{H}^n)$ is isomorphic to $O^+(n, 1) < \mathrm{GL}(n + 1, \mathbb{C})$, so Selberg's Lemma 1.4.19 applies to Γ, and it furnishes a finite-index torsion-free normal subgroup $\Gamma' \lhd \Gamma$. This subgroup acts freely by Proposition 3.1.10. $\qquad\square$

Every finitely generated discrete group $\Gamma < \mathrm{Isom}(\mathbb{H}^n)$ contains at least one finite-index subgroup that acts freely. But how many such subgroups does it contain? Quite a lot, in fact.

Proposition 3.1.15. *Every finitely generated discrete subgroup $\Gamma < \mathrm{Isom}(\mathbb{H}^n)$ is residually finite.*

Proof. See Lemma 1.4.20. $\qquad\square$

The following corollary shows that there is an abundance of torsion-free subgroups. Although algebraic in nature, it has some remarkable geometric consequences, that will be revealed soon in Section 4.3.5.

Corollary 3.1.16. *Let $\Gamma < \mathrm{Isom}(\mathbb{H}^n)$ be discrete and finitely generated. For every non-trivial $g \in \Gamma$ there is a finite-index normal subgroup $\Gamma' \lhd \Gamma$ that acts freely on $\mathbb{H}^n$ and does not contain g.*

3.2. Polyhedra

A polyhedron in $\mathbb{H}^n$ is a natural geometric object, that may be used to visualise discrete groups in $\mathrm{Isom}(\mathbb{H}^n)$ and hyperbolic manifolds. Polyhedra may sometimes be combined to form some tessellations of the space.

3.2.1. Polyhedra. A *half-space* in $\mathbb{H}^n$ is the closure of one of the two portions of space delimited by a hyperplane. We say that a set of half-spaces is locally finite if their boundary hyperplanes are.

Definition 3.2.1. A n-dimensional *polyhedron* P in $\mathbb{H}^n$ is the intersection of a locally finite set of half-spaces. We also assume that P has non-empty interior.

A subset $S \subset \overline{\mathbb{H}^n}$ is *convex* if $x, y \in S$ implies that the segment connecting x, y is also contained in S (such a segment is a half-line or a line if one or both points lie in $\partial \mathbb{H}^n$). Every polyhedron P is clearly convex because it is the intersection of convex sets. Its closure $\overline{P}$ in $\overline{\mathbb{H}^n}$ is also convex.

Let $H \subset \mathbb{H}^n$ be a half-space containing the polyhedron P. If non-empty, the intersection $F = \partial H \cap P$ is called a *face* of P. The *supporting subspace* of F is the smallest subspace of $\mathbb{H}^n$ containing F; the *dimension* of a face is the dimension of its supporting subspace. A face of dimension 0, 1, and $n-1$ is called a *vertex*, an *edge*, and a *facet*.

Exercise 3.2.2. If non-empty, the intersection of faces of P is a face.

Exercise 3.2.3. Every k-dimensional face is a polyhedron in its supporting k-dimensional space.

The *convex hull* of a set $S \subset \overline{\mathbb{H}^n}$ is the intersection of all the convex sets containing S.

Exercise 3.2.4. The convex hull of finitely many points in $\mathbb{H}^n$ that are not contained in a hyperplane is a compact polyhedron. Conversely, every compact polyhedron has finitely many vertices and is the convex hull of them.

Everything we said holds with no modifications for $\mathbb{R}^n$. On S^n some care should be taken: some definitions need to be modified slightly to take into account the annoying presence of antipodal points. We gloss over this technical point.

3.2.2. Finite polyhedra. We now enlarge slightly the class of compact polyhedra by admitting finitely many vertices at infinity.

Definition 3.2.5. A *finite polyhedron* is the convex hull of finitely many points $x_1, \ldots, x_k \in \overline{\mathbb{H}^n}$ that are not contained in the closure of a hyperplane.

The x_i's that lie in $\partial \mathbb{H}^n$ are called *ideal vertices*, while the usual vertices of P are the *finite* or *actual vertices*. The ideal vertices form the set $\overline{P} \setminus P$.
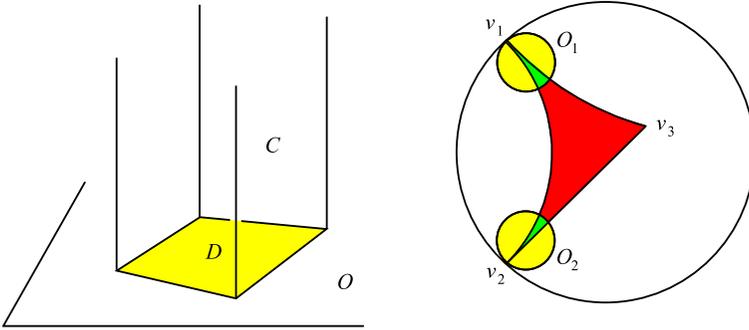
Figure 3.1. The cone C over a domain $D \subset O$ has volume proportional to the area of D (left). If the domain is compact, the cone has finite volume: therefore a finite polyhedron has finite volume (right).

Exercise 3.2.6. A finite polyhedron P has finitely many faces and is the convex hull of its ideal and finite vertices.

We want to estimate the volume of finite polyhedra, and to do this we need a lemma. Given a horosphere O centred at $p \in \partial \mathbb{H}^n$ and a domain $D \subset O$, the *cone* C of D over p is the union of all the half-lines exiting from D and pointing towards p, see Figure 3.1.

Lemma 3.2.7. *Let O be a horosphere centred at $p \in \partial \mathbb{H}^n$, $D \subset O$ any domain and C the cone over D. The following equality holds:*

$$\text{Vol}(C) = \frac{\text{Vol}_O(D)}{n-1}$$

where Vol_O is the $(n-1)$-dimensional volume in the $(n-1)$-manifold O.

Proof. Let O have some height $x_n = h$ as in Figure 3.1. We get

$$\text{Vol}(C) = \int_D dx \int_h^\infty \frac{1}{t^n} dt = \frac{1}{n-1} \int_D \frac{dx}{h^{n-1}} = \frac{1}{n-1} \cdot \text{Vol}_O(D).$$

The proof is complete.   □

We now turn to finite polyhedra.

Proposition 3.2.8. *Every finite polyhedron has finite volume.*

Proof. For every ideal vertex of P, a small horoball centred at p intersects P into a cone that has finite volume. The polyhedron P decomposes into finitely many cones and a bounded region, see Figure 3.1-(right). □

A finite polyhedron without finite vertices is called an *ideal polyhedron*.
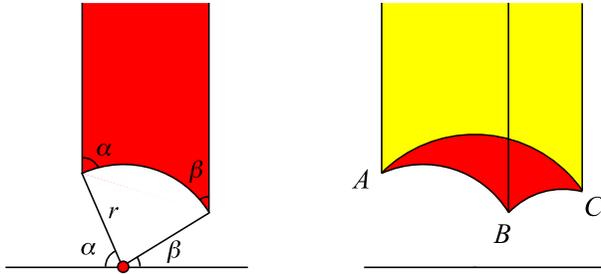
Figure 3.2. A triangle with at least an ideal vertex (left). The area of a triangle with finite vertices can be derived as the area difference of triangles with one ideal vertex (right).

3.2.3. Polygons. A polygon is just a polyhedron of dimension two. In contrast with Euclidean geometry, a strikingly simple formula relates the area of a finite polygon with its inner angles. We define the inner angle of an ideal vertex to be zero.

Proposition 3.2.9. *A polygon P with inner angles $\alpha_1, \ldots, \alpha_n$ has area*

$$\mathrm{Area}(P) = (n-2)\pi - \sum_{i=1}^{n} \alpha_i.$$

Proof. Every polygon decomposes into triangles, and it suffices to prove the formula on these. Consider first a triangle T with at least one vertex at infinity. We use the half-plane model and send this vertex to ∞ as in Figure 3.2-(left). We suppose that the red dot is the origin of $\mathbb{R}^2$, so

$$T = \left\{ (r\cos\theta, y) \mid \beta \leqslant \theta \leqslant \pi - \alpha, \ y \geqslant r\sin\theta \right\}$$

and we get

$$\mathrm{Area}(T) = \int_T \frac{1}{y^2} dx dy = \int_{\pi-\alpha}^{\beta} \int_{r\sin\theta}^{\infty} \frac{-r\sin\theta}{y^2} dy d\theta$$

$$= \int_{\pi-\alpha}^{\beta} -r\sin\theta \left[-\frac{1}{y} \right]_{r\sin\theta}^{\infty} d\theta = \int_{\beta}^{\pi-\alpha} \frac{r\sin\theta}{r\sin\theta} d\theta$$

$$= \int_{\beta}^{\pi-\alpha} 1 = \pi - \alpha - \beta.$$

The area of a triangle with finite vertices ABC is deduced as in Figure 3.2-(right) using the formula

$$\mathrm{Area}(ABC) = \mathrm{Area}(AB\infty) + \mathrm{Area}(BC\infty) - \mathrm{Area}(AC\infty).$$

The proof is complete.   □

The sum of the inner angles of a hyperbolic polygon is strictly smaller than that of a Euclidean polygon with the same number of sides, and the difference between these two numbers is precisely its area.

Figure 3.3. The regular icosahedron and dodecahedron.

Corollary 3.2.10. *Every ideal triangle has area π.*

3.2.4. Platonic solids. The theory of three-dimensional polyhedra in $\mathbb{H}^3$ is very rich: we limit ourselves to the study of the platonic solids.
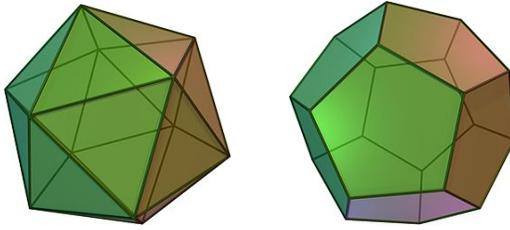
The five Euclidean platonic solids are the regular tetrahedron, the cube, the regular octahedron, icosahedron, and dodecahedron. We now see that each platonic solid $P \subset \mathbb{R}^3$ generates a nice continuous family of solids in the three geometries $\mathbb{H}^3, \mathbb{R}^3$, and S^3.

To construct this family we fix any point x in $\mathbb{H}^3$ and represent P centred at x with varying size. To do this, we put P inside the Euclidean tangent space $T_x \mathbb{H}^3$ centred at the origin and with some radius $t > 0$. Consider the image of its vertices by the exponential map and take their convex hull. We indicate by $P(-t)$ the resulting polyhedron in $\mathbb{H}^3$. The polyhedron $P(-t)$ is combinatorially equivalent to P and has the same symmetries of P.

We extend this family to the other geometries as follows. The polyhedron $P(0)$ is the Euclidean P (unique up to dilations), and $P(t)$ with $t > 0$ is the spherical P, constructed as above with S^3 instead of $\mathbb{H}^3$: we fix $x \in S^3$, take a copy of P inside $T_x S^3$ with radius t, project its vertices, and take the convex hull. We define the spherical $P(t)$ only for $t \in (0, \frac{\pi}{2}]$: when $t = \frac{\pi}{2}$ it degenerates to a hemisphere. We also define $P(-\infty)$ as the ideal hyperbolic platonic solid obtained by sending all the vertices at infinity.

We have defined a polyhedron $P(t)$ for all $t \in [-\infty, \frac{\pi}{2}]$, that lies in $\mathbb{H}^3, \mathbb{R}^3, S^3$ depending on whether t is negative, null, or positive. In some sense the polyhedron $P(t)$ depends continuously on t also when it crosses the value $t = 0$: when $t \to 0$ the polyhedron in $\mathbb{H}^3$ or S^3 shrinks, and every polyhedron tends to a Euclidean one when shrunk.

In particular the dihedral angle $\theta(t)$ of $P(t)$ varies continuously with $t \in [-\infty, \frac{\pi}{2}]$. The function $\theta(t)$ is strictly monotone and we now determine its image. The *vertex valence* of P is the number of edges at each vertex.

polyhedron	$\theta = \frac{\pi}{3}$	$\theta = \frac{2\pi}{5}$	$\theta = \frac{\pi}{2}$	$\theta = \frac{2\pi}{3}$
tetrahedron	ideal $\mathbb{H}^3$	S^3	S^3	S^3
cube	ideal $\mathbb{H}^3$	$\mathbb{H}^3$	$\mathbb{R}^3$	S^3
octahedron			ideal $\mathbb{H}^3$	S^3
icosahedron				$\mathbb{H}^3$
dodecahedron	ideal $\mathbb{H}^3$	$\mathbb{H}^3$	$\mathbb{H}^3$	S^3

Table 3.1. The platonic solids with dihedral angle θ that divide 2π.

Proposition 3.2.11. *Let P have vertex valence $n \in \{3, 4, 5\}$. Then*

$$\theta\left(\left[-\infty, \frac{\pi}{2}\right]\right) = \left[\frac{n-2}{n}\pi, \pi\right].$$

Proof. Since θ is continuous and monotone increasing, it suffices to show that $\theta(-\infty) = \frac{n-2}{n}\pi$ and $\theta(\frac{\pi}{2}) = \pi$.

By intersecting the ideal polyhedron $P(-\infty)$ with a small horosphere O centred at an ideal vertex v we get a regular n-gon in the Euclidean plane O, with interior angles $\frac{n-2}{n}\pi$. The dihedral angle at an edge e incident to v is measured by intersecting $P(-\infty)$ with any hypersurface orthogonal to e: since O is orthogonal to e we get $\theta(-\infty) = \frac{n-2}{n}\pi$.

The polyhedron $P(\frac{\pi}{2})$ is a hemisphere and hence $\theta(\frac{\pi}{2}) = \pi$. $\square$

For some values of t, the platonic solid $P(t)$ may have nice dihedral angles $\theta(t)$, for instance angles that divide 2π. In the Euclidean world, the only platonic solid with such nice dihedral angles is the cube. In the hyperbolic and spherical world we find more.

The platonic solids with dihedral angles that divide 2π are listed in Table 3.1. The table is just a consequence of Proposition 3.2.11: it suffices to know the Euclidean dihedral angles for P, and all the angles bigger (smaller) than this value are spherical (hyperbolic).

In particular, there are four right-angled platonic solids: the spherical tetrahedron, the Euclidean cube, the hyperbolic dodecahedron, and the ideal hyperbolic octahedron.

3.3. Tessellations

A tessellation is a nice paving of $\mathbb{H}^n$ made of polyhedra. Not only tessellations are beautiful objects, but they are also tightly connected with discrete groups of $\mathsf{Isom}(\mathbb{H}^n)$ and hence with hyperbolic manifolds.

Definition 3.3.1. A *tessellation* of $\mathbb{H}^n$ (or $\mathbb{R}^n$, S^n) is a locally finite set of polyhedra that cover the space and intersect only in common faces.

Some examples in spherical and Euclidean space are shown in Figure 3.4. We now construct some families explicitly.
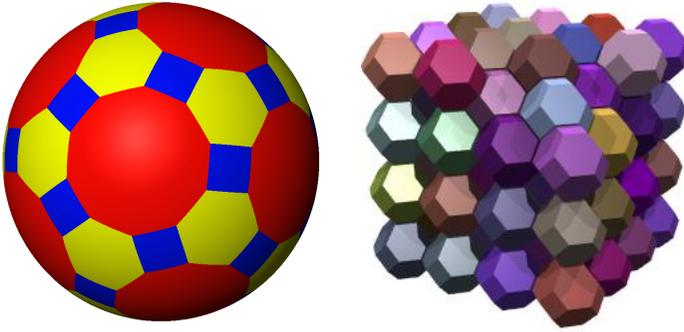
Figure 3.4. A tessellation of S^2 into squares, hexagons, and decagons, and a tessellation of $\mathbb{R}^3$ into truncated octahedra.

Figure 3.5. The tessellations $(2, 3, 3)$, $(2, 3, 4)$ and $(2, 3, 5)$ of the sphere.

3.3.1. Triangles. We want to define some nice tessellations of $\mathbb{H}^2$, $\mathbb{R}^2$, and S^2 into triangles. The following exercise says that every triple of acute angles is realised by some triangle in the appropriate geometry.

Exercise 3.3.2. Given three real numbers $0 < \alpha, \beta, \gamma \leqslant \frac{\pi}{2}$ there is a triangle Δ with inner angles α, β, γ inside $\mathbb{H}^2$, $\mathbb{R}^2$, or S^2 depending on whether the sum $\alpha + \beta + \gamma$ is smaller, equal, or bigger than π.

Let $a, b, c \geqslant 2$ be three natural numbers and Δ be a triangle with inner angles $\frac{\pi}{a}, \frac{\pi}{b}, \frac{\pi}{c}$. The triangle Δ lies in $\mathbb{H}^2$, $\mathbb{R}^2$, or S^2 depending on whether $\frac{1}{a} + \frac{1}{b} + \frac{1}{c}$ is smaller, equal, or bigger than 1. In all cases, by mirroring iteratively Δ along its edges we construct a tessellation of the space.

The triples realisable in S^2 are $(2, 2, c)$, $(2, 3, 3)$, $(2, 3, 4)$, and $(2, 3, 5)$: the last three tessellations are shown in Figure 3.5 and are connected to the platonic solids. They consist of 24, 48, and 120 triangles.

The triples realisable in $\mathbb{R}^2$ are $(2, 3, 6)$, $(2, 4, 4)$, and $(3, 3, 3)$: the tessellations are shown in Figure 3.6. There are infinitely many triples realisable in $\mathbb{H}^2$, and some are shown in Figure 3.7.
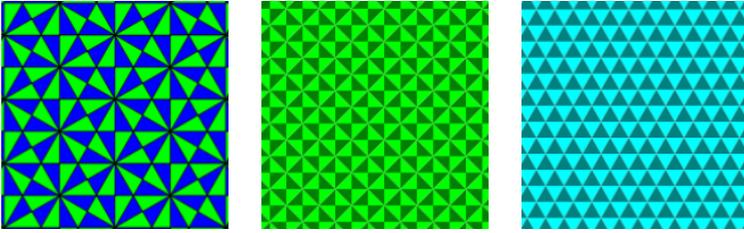
Figure 3.6. The tessellations $(2, 3, 6)$, $(2, 4, 4)$, and $(3, 3, 3)$ of the Euclidean plane.
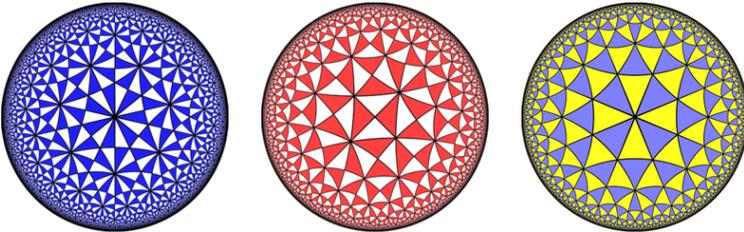
Figure 3.7. The tessellations $(2, 3, 7)$, $(2, 4, 5)$, and $(3, 3, 4)$ of the hyperbolic plane.

In the hyperbolic plane we can also use triangles with vertices at infinity, which have inner angle zero by definition.

Exercise 3.3.3. Given three real numbers $0 \leqslant \alpha, \beta, \gamma \leqslant \frac{\pi}{2}$ with sum smaller than π there is a triangle $\Delta \subset \mathbb{H}^2$ with inner angles α, β, γ.

For any triple (a, b, c) of numbers in $\mathbb{N} \cup \{\infty\}$ with $\frac{1}{a} + \frac{1}{b} + \frac{1}{c} < 1$ we may take the triangle $\Delta \subset \mathbb{H}^2$ with inner angles $\frac{\pi}{a}$, $\frac{\pi}{b}$, $\frac{\pi}{c}$ and reflect it iteratively to get a tessellation of $\mathbb{H}^2$.

The triple (∞, ∞, ∞) gives a nice tessellation into ideal triangles called the *Farey tessellation* and shown in Figure 3.8.

3.3.2. Platonic solids. We now turn to tessellations of 3-dimensional spaces. Table 3.1 displays a finite list of platonic solids P with dihedral angles $\frac{2\pi}{k}$ contained in $\mathbb{H}^3$, $\mathbb{R}^3$, or S^3. For each solid in the list, by reflecting iteratively P along its faces we get a tessellation of the space.

The Figures 3.9, 3.10, and 3.11 show the tessellations of $\mathbb{H}^3$, $\mathbb{H}^3$, and S^3 by regular dodecahedra with dihedral angles $\frac{\pi}{2}$, $\frac{\pi}{3}$, and $\frac{2\pi}{3}$. The first two tessellations contain infinitely many polyhedra and appear in the figures as seen by an observer floating inside $\mathbb{H}^3$. The reader is invited to observe the differences between the first two tessellations of the hyperbolic space. The third tessellation contains finitely many polyhedra and the figure shows its stereographic projection in $\mathbb{R}^3$.
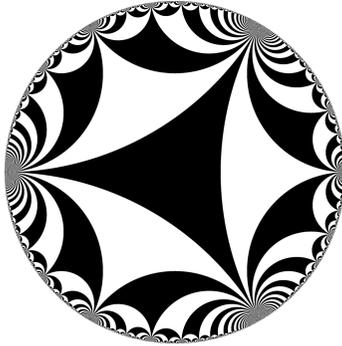
Figure 3.8. The Farey tessellation (∞, ∞, ∞) of the hyperbolic plane.
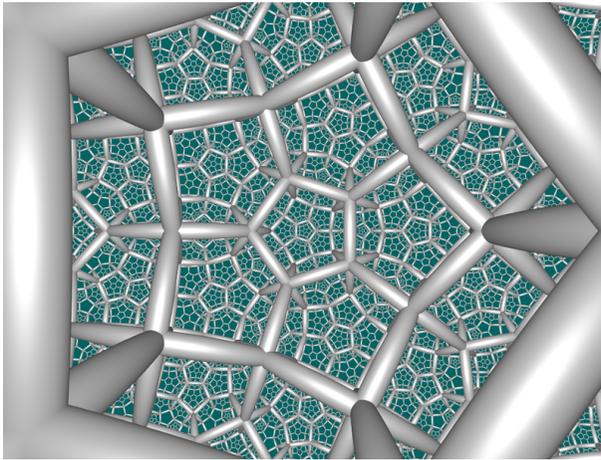
Figure 3.9. The tessellation of $\mathbb{H}^3$ into right-angled regular dodecahedra, seen from inside $\mathbb{H}^3$.

3.3.3. Regular tessellations. Platonic solids are *regular*, that is they have many symmetries. A notion of regularity may be defined in all dimensions for polyhedra and tessellations as follows.

Every tessellation in $\mathbb{H}^n$, $\mathbb{R}^n$, or S^n has a symmetry group, consisting of all the isometries of the ambient space that preserve it. A *flag* is a sequence of faces $f_0 \subset \ldots \subset f_n$ of the tessellation with $\dim f_i = i$. A tessellation is *regular* if the symmetry group acts transitively on flags.
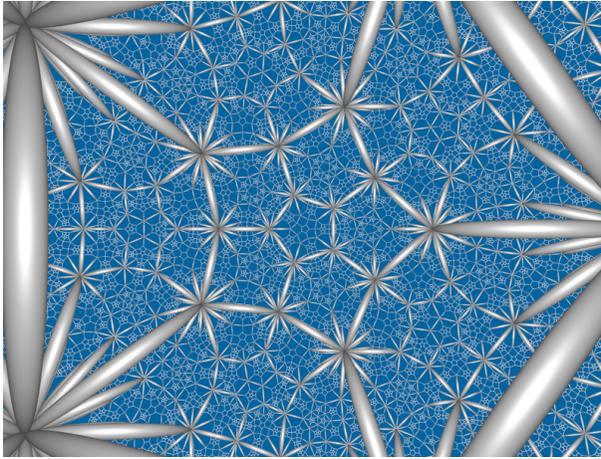
Figure 3.10. The tessellation of $\mathbb{H}^3$ into ideal regular do-
decahedra with dihedral angle $\frac{\pi}{3}$, seen from inside $\mathbb{H}^3$. Note
that all vertices lie at infinity.

These definitions apply also to polyhedra. We note that a regular
polyhedron in $\mathbb{R}^n$, considered up to similarities, may be interpreted as a
regular tessellation in S^{n-1}, and vice versa.

The *Schläfli notation* beautifully encodes various regular tessellations,
in all dimensions and in all geometries. The symbol $\{n\}$ with $n \geqslant 3$ denotes
the regular n-gon, that is a regular tessellation of S^1 by n congruent arcs,
and by extension $\{\infty\}$ denotes the regular tessellation of $\mathbb{R}$ by infinitely
many congruent segments.

The symbol $\{p, q\}$ denotes a regular tessellation in $\mathbb{H}^2, \mathbb{R}^2$, or S^2 by
p-gons where q of them meet at every vertex. The five platonic solids are

$$\{3, 3\}, \{3, 4\}, \{3, 5\}, \{4, 3\}, \{5, 3\},$$

the symbols $\{3, 6\}$, $\{4, 4\}$, and $\{6, 3\}$ denote the two regular tessellations
of $\mathbb{R}^2$ into equilateral triangles, squares, and regular hexagons, and every
other pair $\{p, q\}$ denotes a tessellation of regular p-gons in $\mathbb{H}^2$ with angles
$\frac{2\pi}{q}$, that meet at q at every vertex. We can also interpret $\{p, \infty\}$ as a
regular tessellation of $\mathbb{H}^2$ into ideal regular p-gons (the Farey tessellation
is $\{3, \infty\}$) and $\{\infty, q\}$ as a regular tessellation of ∞-gons that meet at q
at every vertex, whose edges form a q-regular tree (a tree where q edges
meet at every vertex). Finally $\{\infty, \infty\}$ is a regular tessellation of ideal
∞-gons. See Figure 3.12

The symbol $\{p, q, r\}$ denotes a regular tessellation in $\mathbb{H}^3, \mathbb{R}^3$, or S^3 by
polyhedra $\{p, q\}$ where r of them meet at every edge. We deduce from
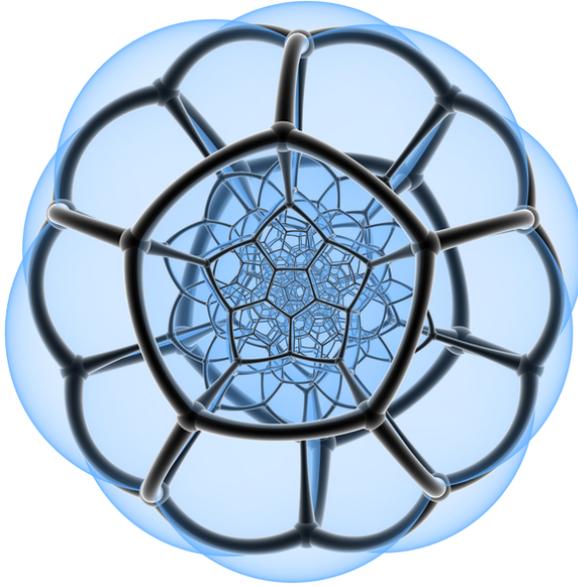
Figure 3.11. The tessellation of S^3 into 120 regular dodec-
ahedra with dihedral angle $\frac{2\pi}{3}$, transposed into $\mathbb{R}^3$ via the
stereographic projection, that transforms straight faces into
round ones but preserves the angles.
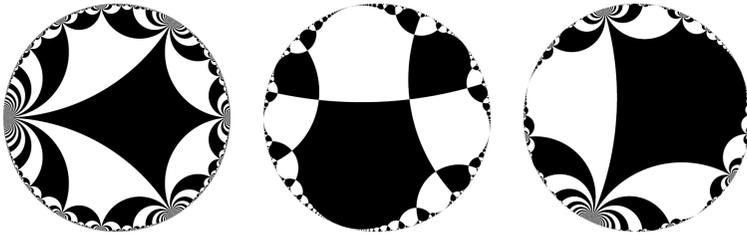
Figure 3.12. The tessellations of $\mathbb{H}^2$ with Schläfli symbols
$\{4, \infty\}$, $\{\infty, 4\}$, and $\{\infty, \infty\}$.

Table 3.1 that the regular tessellations of S^3 are

$$\{3, 3, 3\}, \{3, 3, 4\}, \{3, 3, 5\}, \{3, 4, 3\}, \{4, 3, 3\}, \{5, 3, 3\}.$$

These are the six regular polyhedra (usually called *polytopes*) in dimension
four: we will encounter them again in Chapter 12.2.3. The symbol $\{4, 3, 4\}$
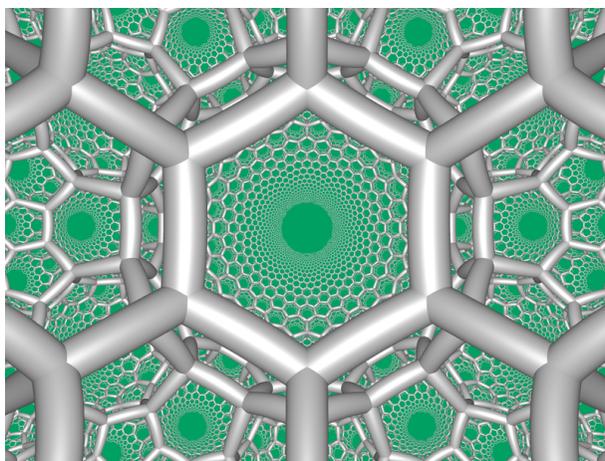
Figure 3.13. In the tessellation $\{6, 3, 3\}$ of $\mathbb{H}^3$ every poly-
hedron has infinitely many hexagonal faces, whose vertices
all lie in a single horosphere.

indicates the tessellation of $\mathbb{R}^3$ into cubes, and

$$\{3, 3, 6\}, \{3, 4, 4\}, \{3, 5, 3\}, \{4, 3, 5\}, \{4, 3, 6\}, \{5, 3, 4\}, \{5, 3, 5\}, \{5, 3, 6\}$$

denote the tessellations of $\mathbb{H}^3$ into platonic solids. It is also possible to
interpret more triples $\{p, q, r\}$ in an appropriate way, sometimes by repre-
senting the vertices of the polyhedra as space-like vectors in the hyperboloid
model.

Every regular tessellation has a *dual* regular tessellation obtained by
taking the barycenters of all the cells involved. The dual of $\{a, b, \ldots, z\}$
is $\{z, \ldots, b, a\}$. Duals of ideal tessellations involve infinite polyhedra: for
instance $\{6, 3, 3\}$ is shown in Figure 3.13.

The story continues in four dimensions: we denote by $\{p, q, r, s\}$ a
regular tessellation made of polytopes of type $\{p, q, r\}$ that meet in s at
every codimension-two face. With the same techniques of Section 3.2.4
we can identify the regular polytopes in $\mathbb{H}^4, \mathbb{R}^4$ and S^4 with dihedral angles
that divide 2π, and the corresponding tessellations of the ambient space –
that we are unfortunately unable to see.

3.3.4. Voronoi tessellations. In the previous section we have con-
structed some tessellations by exploiting the symmetries of regular polyhe-
dra: we can probably extend these methods to less symmetric polyhedra,
but how far can we go? How can we construct highly non-regular tessel-
lations?
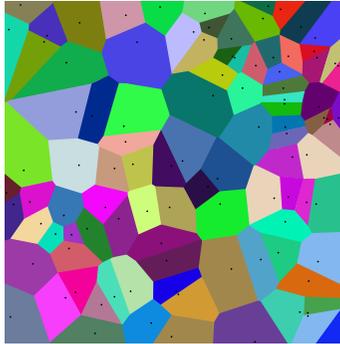
Figure 3.14. A Voronoi tessellation of the Euclidean plane.

There is a strikingly simple procedure that transforms every discrete set S of points in $\mathbb{H}^n$ into a tessellation, called the *Voronoi tessellation* of S. The construction goes as follows.

For every point $p \in S$ we define

$$D(p) = \left\{ q \in \mathbb{H}^n \mid d(q, p) \leqslant d(q, p') \quad \forall p' \in S \right\}.$$

Proposition 3.3.4. *The set $D(p)$ is a polyhedron and the polyhedra $D(p)$ as $p \in S$ varies form a tessellation of $\mathbb{H}^n$.*

Proof. It is an easy exercise to show that the points in $\mathbb{H}^n$ having the same distance from two distinct fixed points form a hyperplane. For every $p' \in S$ different from p we define the half-space

$$H_{p'} = \left\{ q \in \mathbb{H}^n \mid d(q, p) \leqslant d(q, p') \right\}.$$

The set $D(p)$ is the intersection of the half-paces $H_{p'}$ as p' varies in $S \setminus \{p\}$.

Since S is discrete, these half-spaces are locally finite (there are finitely many points in S at bounded distance from p, hence finitely many hyperplanes). Therefore $D(p)$ is a polyhedron. Since S is discrete, every point $q \in \mathbb{H}^n$ has at least one nearest point $p \in S$: therefore the polyhedra $D(p)$ cover $\mathbb{H}^n$ as $p \in S$ varies.

It remains to prove that the polyhedra $D(p)$ intersect along common faces. We have $D(p) \cap D(p') = D(p) \cap \partial H_{p'}$ and hence $D(p) \cap D(p')$ is either empty or a face of $D(p)$. The case of multiple intersections follows from Exercise 3.2.2.   □

Voronoi tessellations of course make sense also in $\mathbb{R}^n$ and S^n, see Figure 3.14. We have just proved that tessellations are not exoteric, but quite ordinary objects, and we now show that they are useful to study hyperbolic manifolds.

3.4. Fundamental domains

We turn back to our discrete subgroups $\Gamma < \mathrm{Isom}(\mathbb{H}^n)$. The geometry of a discrete Γ is nicely controlled by some polyhedra, called *fundamental domains*. We introduce these objects and make some important examples.

3.4.1. Fundamental and Dirichlet domains. Let Γ be a discrete group of isometries of $\mathbb{H}^n$. The group Γ may or may not act freely on $\mathbb{H}^n$.

Definition 3.4.1. A *fundamental domain* for Γ is a polyhedron $D \subset \mathbb{H}^n$ whose translates $g(D)$ as $g \in \Gamma$ varies are distinct and form a tessellation of the space $\mathbb{H}^n$.

If D is a fundamental domain, the group Γ acts freely and transitively on the tessellation $\big\{g(D)\big\}_{g \in \Gamma}$, so in particular the polyhedra $g(D)$ are all isometric.

We describe a procedure that builds a fundamental domain for any discrete subgroup $\Gamma < \mathrm{Isom}(\mathbb{H}^n)$. Pick a point $p \in \mathbb{H}^n$ with trivial stabiliser Γ_p, which exists by Proposition 3.1.9. The group Γ acts freely and transitively on the orbit $\Gamma(p)$, which is discrete by Proposition 3.1.8.

The discrete orbit $\Gamma(p)$ defines a Γ-invariant Voronoi tessellation of $\mathbb{H}^n$, and every polyhedron of the tessellation is a fundamental domain. The polyhedron $D(p)$ of the tessellation containing p is called the *Dirichlet domain* for Γ centred at p. By construction we have

$$D(g(p)) = g(D(p))$$

for all $g \in \Gamma$. We have proved in particular the following.

Proposition 3.4.2. *Every discrete group $\Gamma < \mathrm{Isom}(\mathbb{H}^n)$ has a fundamental domain.*

Fundamental domains are far from being unique. The Dirichlet domain $D(p)$ depends on p in a continuous fashion, and many fundamental domains are not Dirichlet domains.

Exercise 3.4.3. Prove that the shadowed triangle in Figure 3.15 is a fundamental domain for the modular group $\mathbb{PSL}_2(\mathbb{Z})$.

Everything we say also holds for $\mathbb{R}^n$ and S^n.

3.4.2. Fundamental domain of manifolds. Let $\Gamma < \mathrm{Isom}(\mathbb{H}^n)$ be a discrete subgroup that acts freely on $\mathbb{H}^n$ and D be a fundamental domain for Γ. We can get some information on the hyperbolic manifold $M = \mathbb{H}^n/_\Gamma$ by looking at D.

Proposition 3.4.4. *Let $M = \mathbb{H}^n/_\Gamma$ be a hyperbolic manifold and D a fundamental domain for Γ. The projection $\pi \colon \mathbb{H}^n \to M$ restricts to a surjective map $D \to M$ that sends $\mathrm{int}(D)$ isometrically onto an open dense subset of M. In particular we have*

$$\mathrm{Vol}(D) = \mathrm{Vol}(M).$$
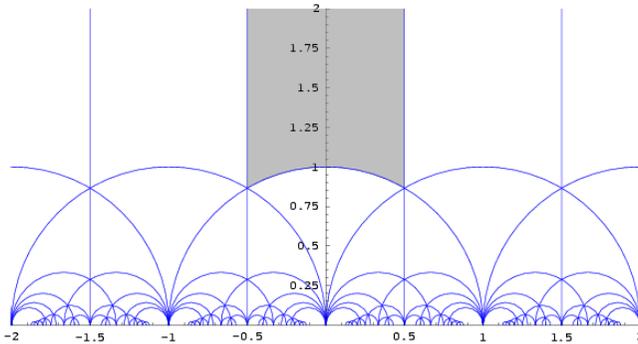
Figure 3.15. The shadowed triangle (with one ideal vertex) is a fundamental domain for $\Gamma = \mathbb{PSL}_2(\mathbb{Z})$ acting on the half-space H^2. The translates of the fundamental domain form the tessellation shown.

If D is a Dirichlet domain, it is compact if and only if M is.

Proof. The translates $g(D)$ cover $\mathbb{H}^n$, hence the projection $D \to M$ is surjective. The translates $g(\mathrm{int}(D))$ are disjoint, hence $\mathrm{int}(D)$ is sent isometrically inside M. The boundary ∂D has measure zero, hence $\mathrm{Vol}(D) = \mathrm{Vol}(\mathrm{int}(D)) = \mathrm{Vol}(M)$.

If D is compact then M clearly is. If $D = D(p)$ is a Dirichlet domain and M is compact, M has finite diameter δ and hence every point in $\mathbb{H}^n$ is at distance $\leqslant \delta$ of some point in the orbit of p. Hence $D(p)$ is contained in the closure of $B(p, \delta)$. $\qquad\square$

A fundamental domain D alone however does not determine M: we will soon see in Section 3.4.8 that a square in $\mathbb{R}^2$ is the fundamental domain of two non homeomorphic compact flat surfaces.

3.4.3. Asteroids. The facets of a fundamental domain D are naturally partitioned into isometric pairs, as follows. Ever facet f of D is incident to D and to another fundamental domain $g(D)$ of the tessellation. The isometry g^{-1} sends $g(D)$ to D and hence sends f to another facet f' of D. One checks immediately that $(f')' = f$, so f and f' are paired isometrically.

One should think of M as obtained from D by identifying these facets in pairs: we can picture an observer − say, a spaceship − floating and traveling inside M by visualising it in D, and jumping from f to f' every time it crosses the interior of a facet f, like in the 1979 video game *Asteroids*.

3.4.4. Spine and cut locus. Another picture that may help to understand M geometrically is the *spine* $S \subset M$ defined as $S = \pi(\partial D)$. This is a $(n-1)$-dimensional object in M whose complement is an open ball.

When $D = D(p)$ is the Dirichlet domain of a point p, the spine S is called the *cut locus* of the point $q = \pi(p) \in M$.

Exercise 3.4.5. The cut locus of $q \in M$ consists of all points q' such that there are at least two geodesics of minimal length connecting q to q'.

We now exhibit some important examples of discrete groups of isometries in the three geometries $\mathbb{H}^n$, $\mathbb{R}^n$, and S^n, and study their fundamental domains.

3.4.5. Triangle groups. Let $a, b, c \geqslant 2$ be natural numbers and Δ be the triangle in $\mathbb{H}^2$, S^2, or $\mathbb{R}^2$ with inner angles $\frac{\pi}{a}$, $\frac{\pi}{b}$, $\frac{\pi}{c}$. By reflecting iteratively Δ along its sides we get a tessellation T, see Section 3.3.1.

The *triangle group* $\Gamma = \Gamma(a, b, c)$ is the group of isometries of S^2, $\mathbb{R}^2$, or $\mathbb{H}^2$ generated by the reflections x, y, z along the three sides of Δ opposite to the vertices with inner angles a, b, c, respectively.

Proposition 3.4.6. *The triangle group $\Gamma(a, b, c)$ acts freely and transitively on the triangles of the tessellation T. Hence it is discrete and Δ is a fundamental domain for Γ. A presentation for the group is*

$$\langle x, y, z \mid x^2, y^2, z^2, (xy)^c, (yz)^a, (zx)^b \rangle.$$

Proof. We restrict for simplicity to the hyperbolic case, the others being analogous. It is convenient to construct the tessellation T abstractly (this also furnishes a rigorous proof that by mirroring Δ along its sides we get a tessellation of $\mathbb{H}^2$).

We denote by x, y, z both the sides of Δ and the reflections along them. For every $g \in \Gamma$ we define an abstract copy Δ_g of Δ, and then we glue all these abstract copies altogether by pairing their sides as follows: for every side of Δ, say x, and every $g \in \Gamma$, we identify the two copies of the side x in Δ_g and Δ_{gx} with the obvious identity map.

Via these identifications we get an abstract space T tessellated into the triangles Δ_g. The group Γ acts freely and transitively on the tessellation (the element $g' \in \Gamma$ sends Δ_g to $\Delta_{g'g}$). We now prove that T has a natural structure of a hyperbolic surface: since Γ acts transitively, it suffices to check this for the points p lying in Δ_e. If p lies in the interior of Δ_e or of a side, this is clear by construction. If p is a vertex, say with inner angle a, by construction a cycle of $2a$ triangles $\Delta_e, \Delta_y, \Delta_{yz}, \Delta_{yzy}, \Delta_{yzyz}, \ldots$ is attached around p because $(yz)^a = e$, and hence T is naturally a hyperbolic surface also near p.

The hyperbolic surface T is easily seen to be connected and complete. There is a natural developing map $\varphi \colon T \to \mathbb{H}^2$ that sends Δ_g to $g(\Delta)$. The map φ is a local isometry, hence a covering by Proposition 1.2.19, hence an isometry since $\mathbb{H}^2$ is simply connected. We identify T with $\mathbb{H}^2$ via φ.

It remains to prove that Γ may be presented as stated. Let G be the group presented as

$$\langle x, y, z \mid x^2, y^2, z^2, (xy)^c, (yz)^a, (zx)^b \rangle.$$

There is a natural surjection $G \to \Gamma$. We may repeat all the arguments above using G instead of Γ and get another T' isometric to $\mathbb{H}^2$ that covers T. We deduce that $T' = T$ and $G = \Gamma$. □

The triangle group $\Gamma(2,2,c) \cong \mathbb{Z}/_{2\mathbb{Z}} \times D_{2c}$ has order $4c$ and is the symmetry group of a prism. The triangle groups $\Gamma(2,3,3)$, $\Gamma(2,3,4)$, and $\Gamma(2,3,5)$ have order 24, 48, and 120 and are the symmetry groups of the regular tetrahedron, octahedron, and icosahedron: the reader is invited to check all these facts visually by looking at Figure 3.5.

The triangle groups $\Gamma(2,3,6)$, $\Gamma(2,4,4)$, and $\Gamma(3,3,3)$ are discrete groups of isometries of $\mathbb{R}^2$ with compact quotient $\mathbb{R}^2/_{\Gamma}$. If $\frac{1}{a} + \frac{1}{b} + \frac{1}{c} < 1$ the group $\Gamma(a,b,c)$ is a discrete subgroup of $\mathrm{Isom}(\mathbb{H}^2)$. It contains infinitely many elliptic elements, such as reflections along lines and finite-order rotations around the vertices of the triangles. By Selberg's Lemma, there is a torsion-free subgroup $\Gamma' < \Gamma$ of some finite index h. The quotient $\mathbb{H}^2/_{\Gamma'}$ is a closed hyperbolic surface and is tessellated into h triangles isometric to Δ.

3.4.6. Coxeter polyhedra. Triangle groups may be generalised to all dimensions as follows. A polyhedron P in $\mathbb{H}^n$ (or $\mathbb{R}^n$, S^n) is a *Coxeter polyhedron* if the dihedral angle of every codimension-two face divides π. For instance, the regular ideal tetrahedron and octahedron are Coxeter polyhedra.

The following theorem generalises Proposition 3.4.6. Let P be a finite Coxeter polyhedron: it is the convex hull of finitely many vertices in $\overline{\mathbb{H}^n}$ (or $\mathbb{R}^n$, S^n) and has k facets, that we number as $1, \dots, k$; we denote by r_i the reflection along the i-th facet, and by $\frac{\pi}{a_{ij}}$ the dihedral angle formed by the i-th and j-th facets, if they are incident. Let Γ be the isometry group generated by the reflections along the facets of P.

Theorem 3.4.7. *By mirroring iteratively a finite Coxeter polyhedron P along its facets we get a tessellation of $\mathbb{H}^n$ (or $\mathbb{R}^n$, S^n). The group Γ acts freely and transitively on the tessellation: hence it is discrete and P is a fundamental domain for Γ. A presentation for Γ is*

$$\langle r_1, \dots, r_k \mid r_i^2, (r_i r_j)^{a_{ij}} \rangle$$

where i varies in $1, \dots, k$ and the pair i, j varies among the incident facets.

Proof. Same proof as Proposition 3.4.6, with in addition an induction on the dimension n. Here are the details. We consider for simplicity only the hyperbolic case.

For every $g \in \Gamma$ we define an abstract copy P_g of P, and we identify the i-th facet of P_g and P_{gr_i} for all $g \in \Gamma$ and all i. To prove that the resulting space T is naturally a hyperbolic manifold, we use the induction hypothesis as follows. Every $p \in P_e$ lies in the interior of some h-dimensional face f, and let $\Gamma_f < \Gamma$ be the subgroup generated by the reflections along all the facets of P containing f. The point p is adjacent in T to the

polyhedra P_g such that $g \in \Gamma_f$. By intersecting each such P_g with a small codimension $h + 1$ sphere centred in p and contained in the codimension h subspace orthogonal to f, we get some spherical polytope Q_g of dimension $n - h - 1 < n$. (The spherical polytope Q_t is usually called the *spherical link* of f.)

By construction $Q_e \subset S^{n-h-1}$ is a Coxeter polytope and the subgroup $\Gamma_f < \mathrm{Isom}(S^{n-h-1})$ is generated by the reflections along its facets. By applying the induction hypothesis on Q_e we deduce that the polytopes Q_g with $g \in \Gamma_f$ form a tessellation of S^{n-h-1}, hence the polyhedra P_g incident to p form naturally a hyperbolic ball locally near p. Therefore T is a hyperbolic manifold.

Proving that T is complete requires a bit of care if P_e has some ideal vertex v. In that case v has a Euclidean link (obtained by intersecting P_e with a small horosphere) and we conclude by induction as above that the Euclidean links of the P_g incident to v in T glue to form a Euclidean $\mathbb{R}^{n-1}$, that is a horosphere, so by intersecting each P_g with small horoballs we get a horoball near v, which is complete.

The hyperbolic space T is complete and connected, and we conclude as in the proof of Proposition 3.4.6.   □

A group generated by some reflections along hyperplanes is called a *reflection group*. The following proposition shows that Coxeter polyhedra generate all the interesting reflection groups.

Proposition 3.4.8. *Every discrete reflection group* Γ *is generated by the reflections along the facets of some Coxeter polyhedron.*

Proof. Consider the mirror hyperplanes of all the reflections in Γ. Since Γ is discrete, these form a locally finite set and hence define a tessellation of $\mathbb{H}^n$ onto which Γ acts transitively. Pick one polyhedron P of the tessellation. The reflections along the facets of P generate Γ (exercise).   □

Coxeter polyhedra are beautiful objects that can be used to construct hyperbolic manifolds: every finite Coxeter polyhedron P generates a reflection group Γ which contains, by Selberg's Lemma, a torsion-free subgroup Γ' of some finite index h. The quotient $M = \mathbb{H}^n/_{\Gamma'}$ is a hyperbolic manifold and is tessellated into h copies of P, so that $\mathrm{Vol}(M) = h\mathrm{Vol}(P)$. By residual finiteness, there are plenty of such manifolds.

3.4.7. Coxeter graphs. If a Coxeter polyhedron P has a reasonable number of facets (for instance, if P is a simplex), then one can study many of its properties by looking at its *Coxeter graph*, which is constructed as follows: draw one node for each facet of P, and for every pair of facets intersecting with dihedral angle $\frac{\pi}{a}$ connect the corresponding nodes with an edge labeled with a. One also uses thickened or dashed edges to denote non-incident faces that do or do not intersect at some ideal
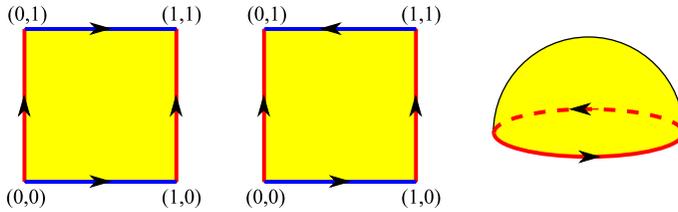
Figure 3.16. A fundamental domain in $\mathbb{R}^2$ for the torus (left) and the Klein bottle (centre): opposite sides should be identified as indicated by the arrows. A fundamental domain in S^2 for $\mathbb{RP}^2$ (right).

vertex, respectively (these cases cannot occur on simplexes). Given the abundance of right angles, one usually omit the edges with label 2.

Exercise 3.4.9. Let P be a regular polytope or tessellation with Schläfli symbol $\{a, b, \ldots, z\}$. By quotienting P via all its isometries we get a Coxeter simplex, called its *characteristic simplex*, whose Coxeter graph is

$$\bullet \overset{a}{\quad} \bullet \overset{b}{\quad} \bullet \quad \cdots \quad \overset{z}{\quad} \bullet$$
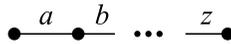
3.4.8. Flat tori. The isometry group $\mathrm{Isom}(\mathbb{R}^n)$ of $\mathbb{R}^n$ contains the translations subgroup that we identify with $\mathbb{R}^n$. The integer translations $\Gamma = \mathbb{Z}^n$ form a discrete subgroup that acts freely on $\mathbb{R}^n$, therefore the quotient $\mathbb{R}^n/_\Gamma$ is a flat manifold. It is naturally diffeomorphic to the n-dimensional torus:

$$\mathbb{R}^n/_{\mathbb{Z}^n} = (\mathbb{R}/_{\mathbb{Z}})^n \cong \underbrace{S^1 \times \ldots \times S^1}_{n}.$$

Exercise 3.4.10. For every $p \in \mathbb{R}^n$ the Dirichlet domain $D(p)$ is a n-dimensional unit cube centred at p.

The flat n-torus may be seen as the unit n-cube with its opposite facets identified by translations. The two-dimensional case is shown in Figure 3.16-(left): by identifying the opposite sides of a square we get a torus.

The n-torus possesses a continuous family of non-isometric flat metrics. A *lattice* $\Gamma < \mathbb{R}^n$ is a discrete subgroup isomorphic to $\mathbb{Z}^n$ which spans $\mathbb{R}^n$ as a vector space. We see Γ as a group of translations.

Exercise 3.4.11. For every lattice Γ, the flat manifold $\mathbb{R}^n/_\Gamma$ is diffeomorphic to the n-torus. A fundamental domain is the parallelotope spanned by n generators of Γ.

A Dirichlet domain for Γ is almost never a parallelotope! In the *hexagonal torus* $\mathbb{R}^2/_\Gamma$ the group Γ is the equilateral lattice generated by the translations

$$(1, 0), \quad \left(\tfrac{1}{2}, \tfrac{\sqrt{3}}{2}\right)$$

and the Dirichlet domain of any point is a regular hexagon. Orientable flat manifolds are easily classified (up to diffeomorphism) in dimension two.

Proposition 3.4.12. *Every closed flat orientable surface is a torus.*

Proof. A closed flat surface S is isometric to $\mathbb{R}^2/_\Gamma$ for some discrete group Γ of orientation-preserving isometries acting freely. Every fixed-point-free orientation-preserving isometry of $\mathbb{R}^2$ is a translation (exercise), hence $\Gamma < \mathbb{R}^2$. If Γ has rank one then S is not compact, hence Γ is a lattice and S is a torus. $\square$

Flat manifolds more complicated than tori can be constructed by considering also non-translational isometries of $\mathbb{R}^n$, which exist for $n \geqslant 3$ and also for $n = 2$ in the non-orientable setting. For instance, in dimension two we find the *Klein bottle* by taking Γ as the group generated by the isometries:

$$\tau\colon (x, y) \mapsto (x + 1, y), \qquad \eta\colon (x, y) \mapsto (1 - x, y + 1).$$

A fundamental domain for the Klein bottle is shown in Figure 3.16-(centre).

Every subgroup $\Gamma' < \Gamma$ furnishes another flat manifold $\mathbb{R}^2/_{\Gamma'}$ that covers the Klein bottle. For instance, the subgroups $\langle \tau \rangle$ and $\langle \eta \rangle$ generated respectively by τ and η are both isomorphic to $\mathbb{Z}$ but provide two different coverings: the manifold $\mathbb{R}^2/_{\langle \tau \rangle}$ is an infinite cylinder while $\mathbb{R}^2/_{\langle \eta \rangle}$ is an infinite Möbius strip.

The subgroup Γ' generated by the translations τ and η^2 is isomorphic to $\mathbb{Z}^2$ and has index 2 in Γ. The Klein bottle is doubly covered by the flat torus $\mathbb{H}^2/_{\Gamma'}$. A fundamental domain for Γ' is a rectangle with vertices $(0, 0)$, $(1, 0)$, $(0, 2)$, $(1, 2)$.

We have seen that the Klein bottle is covered by a torus. In fact, we will see in Section 4.4 that *every* closed flat n-manifold is covered by a flat n-torus.

Exercise 3.4.13. A square $Q \subset \mathbb{R}^2$ is a Coxeter polygon and determines a reflection group Γ. What is the minimum index of a torsion-free subgroup $\Gamma' < \Gamma$? Which topological surfaces $\mathbb{R}^2/_{\Gamma'}$ do we get?

3.4.9. Real projective spaces. We now exhibit some elliptic manifolds. Every elliptic manifold is covered by S^n and is hence compact and has finite fundamental group (because coverings between compact manifolds have finite degree).

An important example is the real projective space $\mathbb{RP}^n = S^n/_\Gamma$ where Γ is the order-two group generated by the antipodal map $\iota(x) = -x$.

Exercise 3.4.14. For every $p \in S^n$, the Dirichlet domain $D(p)$ of Γ is the hemisphere centred at p.

The two-dimensional case is shown in Figure 3.16-(right). The only elliptic surfaces are S^2 and $\mathbb{RP}^2$ in virtue of the following.

Proposition 3.4.15. *In even dimension n the only elliptic manifolds are* S^n *and* $\mathbb{RP}^n$.

Proof. Let $M = S^n/_\Gamma$ be an elliptic manifold. Every matrix in $SO(n+1)$ has an eigenvalue $+1$ because n is even (exercise) and hence acts in S^n with a fixed point. Therefore Γ contains no non-trivial orientation-preserving isometries, and since these form a group of index at most two, either Γ is trivial or is generated by an orientation-reversing fixed-point free involution in $O(n+1)$, and $-I$ is the only such element (exercise). $\qquad\square$

3.4.10. Lens spaces. In dimension two $\mathbb{RP}^2$ and S^2 are the only elliptic manifolds, but in dimension three there are more. Let $p \geqslant 1$ and $q \geqslant 1$ be coprime integers and set $\omega = e^{\frac{2\pi i}{p}}$. We identify $\mathbb{R}^4$ with $\mathbb{C}^2$ and see S^3 as

$$S^3 = \left\{ (z, w) \in \mathbb{C}^2 \mid |z|^2 + |w|^2 = 1 \right\}.$$

The map

$$f(z, w) = (\omega z, \omega^q w)$$

is an isometry of $\mathbb{R}^4$ because it consists of two simultaneous rotations on the coordinate planes $w = 0$ and $z = 0$. The map f hence induces an isometry of S^3. It has order p and none of its iterates $f, f^2, \ldots, f^{p-1}$ has a fixed point. Therefore the group $\Gamma = \langle f \rangle$ generated by f acts freely on S^3, and is discrete because it is finite.

We have constructed an elliptic manifold $S^3/_\Gamma$, called a *lens space* and indicated with the symbol $L(p, q)$. Its fundamental group is isomorphic to $\Gamma = \mathbb{Z}/_{p\mathbb{Z}}$. Note that the manifold depends on both p and q.

3.5. Geodesic boundary, non-complete, and cone manifolds

We now introduce three important variations on the complete hyperbolic manifolds theme. The first consists of admitting a totally geodesic boundary, the second is a brief overlook of some phenomena that may occur when the completeness hypothesis is dropped, and in the third we allow some conical singularity on a codimension-two geodesic stratum.

3.5.1. Hyperbolic manifolds with geodesic boundary. We reformulate a definition of hyperbolic (elliptic, flat) manifolds that allows the presence of some geodesic boundary. These manifolds are useful because they can be glued along their boundaries to produce new hyperbolic (elliptic, flat) manifolds.

Definition 3.5.1. A *hyperbolic* (*elliptic, flat*) *manifold M with geodesic boundary* is a Riemannian manifold with boundary where every point has an open neighbourhood isometric to an open set in a half-space in $\mathbb{H}^n$ (S^n, $\mathbb{R}^n$).

The boundary ∂M of a hyperbolic (elliptic, flat) n-manifold with geodesic boundary is a hyperbolic (elliptic, flat) $(n-1)$-manifold without boundary. Theorem 3.1.2 extends appropriately to this context.
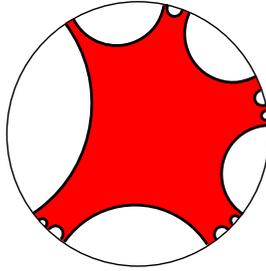
Figure 3.17. An intersection of (possibly infinitely many!) half-planes. The universal cover of a hyperbolic surface with boundary is isometric to such an object.

Theorem 3.5.2. *Every complete simply connected hyperbolic manifold M with geodesic boundary is isometric to the intersection of some half-spaces in $\mathbb{H}^n$ with disjoint boundaries.*

Proof. We construct a developing map $D\colon M \to \mathbb{H}^n$ as in Theorem 3.1.2. The map D is injective, because every two points $p, q \in M$ are connected by a geodesic (exercise: use completeness and geodesic boundary), which is sent to a geodesic via D, so p, q are sent to distinct points $D(p), D(q)$.

Therefore $D(M) \subset \mathbb{H}^n$ is isometric to M. Since it is complete and with geodesic boundary, its boundary consists of disjoint hyperplanes, and hence $D(M)$ is the intersection of half-spaces bounded by them. $\square$

An example is sketched in Figure 3.17. Every complete hyperbolic manifold with geodesic boundary can be enlarged to a hyperbolic manifold without boundary in a canonical way.

Corollary 3.5.3. *Every complete hyperbolic n-manifold M with geodesic boundary is contained in a complete hyperbolic n-manifold N without boundary, such that $N \setminus \mathrm{int}(M)$ is diffeomorphic to $\partial M \times [0, +\infty)$.*

Proof. The proof of Proposition 3.1.3 applies and shows that $M = \tilde{M}/\Gamma$ where $\tilde{M} \subset \tilde{\mathbb{H}}^n$ is an intersection of half-spaces and $\Gamma < \mathrm{Isom}(\tilde{M})$ acts freely and properly discontinuously on $\tilde{M}$.

Every local isometry in $\mathbb{H}^n$ extends to a global isometry, therefore $\Gamma < \mathrm{Isom}(\tilde{M}) < \mathrm{Isom}(\mathbb{H}^n)$. The group Γ acts freely on $\tilde{M}$ and hence also on $\mathbb{H}^n$: if it had a fixed point $x \in \mathbb{H}^n$, it would fix also the unique point $y \in \tilde{M}$ that is the closest to x. The manifold $N = \mathbb{H}^n/\Gamma$ contains naturally M.

For every $p \in \partial M$, let $\gamma_p(t)$ be the unit speed geodesic starting at p orthogonal to ∂M and directed outward. By looking at the universal cover one sees easily that the map $\partial M \times [0, +\infty) \to N \setminus \mathrm{int}(M)$, $(p, t) \mapsto \gamma_p(t)$ is a diffeomorphism (we use here that $\tilde{M} \subset \mathbb{H}^n$ is convex). $\square$

3.5.2. Cut and paste. Hyperbolic manifolds with geodesic boundary are useful because they can be glued to produce new hyperbolic manifolds.

Let M and N be hyperbolic manifolds with geodesic boundary and $\psi \colon \partial M \to \partial N$ be an isometry. Let $M \cup_\psi N$ be the topological space obtained by quotienting the disjoint union $M \sqcup N$ by the equivalence relation that identifies p to $\psi(p)$ for all $p \in \partial M$.

Proposition 3.5.4. *The space $M \cup_\psi N$ has a natural structure of hyperbolic manifold.*

Proof. The interiors of M and N inherit their hyperbolic metrics. When we glue p to $\psi(p)$ we get a point q that has two half-disc neighbourhoods on its sides, which glue to a honest hyperbolic disc, inducing a hyperbolic metric near q. □

Conversely, if an orientable hyperbolic manifold M contains an orientable geodesic hypersurface N, we can cut M along N to get a hyperbolic manifold with geodesic boundary. The boundary will consist of two copies of N.

3.5.3. Non-complete hyperbolic manifolds. There is no classification of simply connected non-complete hyperbolic manifolds: for instance, we may get plenty of uninteresting examples by removing complicated closed sets from $\mathbb{H}^n$. However, the first part of the proof of Theorem 3.1.2 still applies and provides the following:

Proposition 3.5.5. *Let M be a non-complete simply connected hyperbolic n-manifold. There is a local isometry*

$$D \colon M \to \mathbb{H}^n$$

which is unique up to post-composing with $\mathsf{Isom}(\mathbb{H}^n)$.

Proof. Construct D as in the proof of Theorem 3.1.2: the completeness of M is used there only in the last paragraph to show that D is a covering. As a local isometry, the map D is determined by its first-order behaviour at any point $p \in M$, and is hence unique up to post-composing with an isometry of $\mathbb{H}^n$. □

The map D is called a *developing map* and is neither injective nor surjective in general. We can push-forward isometries along D as follows.

Proposition 3.5.6. *Let M be a hyperbolic n-manifold and $D \colon M \to \mathbb{H}^n$ be a local isometry. For every $g \in \mathsf{Isom}(M)$ there is a unique $\rho(g) \in \mathsf{Isom}(\mathbb{H}^n)$ such that*

$$\rho(g) \circ D = D \circ g.$$

The resulting map $\rho \colon \mathsf{Isom}(M) \to \mathsf{Isom}(\mathbb{H}^n)$ is a homomorphism.

Proof. Pick a point $p \in M$ and define $\rho(g)$ as the unique isometry of $\mathbb{H}^n$ such that $\rho(g) \circ D = D \circ g$ on a small neighbourhood of p. This equality between local isometries holds locally and hence globally on M. The map ρ is easily checked to be a homomorphism. $\square$

The homomorphism ρ is the *holonomy* associated to D. If M is a non-complete hyperbolic n-manifold, its universal covering $\pi \colon \tilde{M} \to M$ inherits a non-complete hyperbolic metric. Therefore we get a developing map

$$D \colon \tilde{M} \to \mathbb{H}^n$$

together with a holonomy

$$\rho \colon \mathrm{Aut}(\pi) \longrightarrow \mathrm{Isom}(\mathbb{H}^n)$$

which is the restriction of ρ to the subgroup $\mathrm{Aut}(\pi) < \mathrm{Isom}(\tilde{M})$. By fixing a point in $\tilde{M}$ we identify $\mathrm{Aut}(\pi)$ with $\pi_1(M)$ and get a holonomy

$$\rho \colon \pi_1(M) \longrightarrow \mathrm{Isom}(\mathbb{H}^n).$$

As every metric space, a non-complete Riemannian manifold M has a unique *completion* $\overline{M}$, a complete metric space that contains M as an open dense set. The completion $\overline{M}$ is however not necessarily a manifold, except in some lucky cases. These lucky cases are extremely important in dimension three, as we will see in Chapter 14. In that chapter we will understand the completion $\overline{M}$ of a hyperbolic three-manifold by studying its developing map and holonomy.

3.5.4. Singularities. A particular class of non-complete hyperbolic manifolds deserves our attention, because their completions are some nice and natural objects: manifolds with *cone angles* along some codimension-two singular geodesic stratum.

Let $S \subset \mathbb{H}^n$ be a codimension-two subspace. The incomplete hyperbolic manifold $\mathbb{H}^n \setminus S$ has fundamental group $\mathbb{Z}$ and we denote by

$$\pi \colon X \longrightarrow \mathbb{H}^n \setminus S$$

its universal covering: here X is an interesting non-complete simply connected hyperbolic manifold. Note that π may be interpreted as a developing map $\pi \colon X \to \mathbb{H}^n$ which is neither injective nor surjective.

Exercise 3.5.7. The metric completion of X is $\overline{X} = X \sqcup \tilde{S}$ where $\tilde{S}$ is an identical copy of S. The covering π extends to a surjective map

$$\pi \colon \overline{X} \to \mathbb{H}^n$$

that sends $\tilde{S}$ to S. If $p \in \tilde{S}$ and $q \in \overline{X}$ then $d(p, q) = d(\pi(p), \pi(q))$.

The space $\tilde{S}$ is the *singularity* of $\overline{X}$ and should be interpreted as lying in $\overline{X}$ with infinite cone angle: the space $\overline{X}$ looks like a book with infinitely many pages (it is not locally compact at $\tilde{S}$). Singularities with finite cone angles will be defined in the next section as quotients of this model. The

dimensions that will be interesting for us are of course $n = 2$ and 3, where $\tilde{S}$ is a point and a line, respectively.

3.5.5. Cone manifolds. For every $\theta \in \mathbb{R}$ there is a well-defined rotation $\mathrm{Rot}_\theta \colon \overline{X} \to \overline{X}$ of angle θ around $\tilde{S}$, which projects to a rotation in $\mathbb{H}^n$ of angle θ around S; note that $\mathrm{Rot}_\theta \neq \mathrm{id}$ for all $\theta \neq 0$, including $\theta = 2k\pi$.

We fix $\theta \neq 0$, so that Rot_θ generates a free cyclic group $\Gamma < \mathrm{Isom}(\overline{X})$.

Exercise 3.5.8. The metric space $\overline{X}/\Gamma$ is the completion of X/Γ and is homeomorphic to $\mathbb{R}^n$.

The image of $\tilde{S}$ along the quotient $\pi \colon \overline{X} \to \overline{X}/\Gamma$ is a copy of $\tilde{S}$ and should now be interpreted as a singularity with cone angle θ. When $\theta = 2k\pi$ for some integer k the map $\pi \colon \overline{X} \to \mathbb{H}^n$ factors as

$$\overline{X} \xrightarrow{\pi_1} \overline{X}/\Gamma \xrightarrow{\pi_2} \mathbb{H}^n.$$

When $k = \pm 1$ the map π_2 is an isometry. We introduce the following definition.

Definition 3.5.9. A *hyperbolic manifold M with cone angles* is a metric space with charts in some $\overline{X}/\Gamma$ and transition maps that are isometries.

The term "isometries" should be interpreted in the strongest sense: the transition maps send singular points to singular points, and outside of the singular points these are isometries of Riemannian manifolds.

The points in M that are mapped along charts to some singular set form a hyperbolic codimension-two submanifold in M with some cone angle θ, and different components may have different cone angles. Points with cone angle 2π may be considered as ordinary points, while points with cone angle different from 2π are *singular* and form the *singular locus* of M.

The singular locus of M is a geodesic codimension-two manifold, and its complement is an ordinary hyperbolic manifold, whose metric completion is M. Flat and spherical manifolds with cone angles are defined in the same way.

3.5.6. Examples. The following simple construction is a source of many two-dimensional examples: let P be any polygon in $\mathbb{H}^2, \mathbb{R}^2$, or S^2, with some inner angles $\alpha_1, \ldots, \alpha_k$. By doubling P along its edges we get a topological sphere with k cone points of angles $2\alpha_1, \ldots, 2\alpha_k < 2\pi$. This is a hyperbolic, flat, or spherical surface with cone angles.

Another simple but maybe more intriguing construction is the following: pick any polyhedron P in $\mathbb{H}^3, \mathbb{R}^3$, or $\mathbb{S}^3$. The faces of P are (hyperbolic, flat, or elliptic) polygons that glue isometrically along the edges: hence ∂P is a topological sphere with a natural structure of (hyperbolic, flat, or elliptic) manifold with cone angles at the vertices.

To construct examples in dimension three we need a bit more work because polyhedra have more strata. Non-compact examples are easier to build: let $P \subset \mathbb{H}^3$ be any ideal hyperbolic polyhedron, with some k edges

with dihedral angles $\alpha_1, \ldots, \alpha_k$; the double of P is naturally a hyperbolic cone manifold where the edges form the singular locus with cone angles $2\alpha_1, \ldots, 2\alpha_k$.

3.6. Orbifolds

An orbifold is an object locally modelled on finite quotients of $\mathbb{R}^n$. It naturally arises when we quotient a Riemannian manifold by a discrete group of isometries that may not act freely. Orbifolds behave like manifolds on many aspects.

3.6.1. Definition. Let $\Gamma < O(n)$ be a finite group of linear isometries and $V \subset \mathbb{R}^n$ be a Γ-invariant open set. The resulting map

$$\varphi \colon V \longrightarrow V/_\Gamma$$

is called a *local orbifold model*. For instance, we may pick $V = B(0, r)$. We now generalise manifolds by allowing local orbifold models in the atlas.

Definition 3.6.1. Let O be a Hausdorff paracompact topological space. An *orbifold atlas* on O is an open covering $\{U_i\}_{i \in I}$ of O, closed by finite intersections and equipped with local orbifold models

$$\varphi_i \colon V_i \longrightarrow V_i/_{\Gamma_i} = U_i.$$

The local models are connected by some appropriate transition functions: for every inclusion $U_i \subset U_j$ there is an injective homomorphism

$$f_{ij} \colon \Gamma_i \hookrightarrow \Gamma_j$$

and a Γ_i-equivariant smooth embedding $\psi_{ij} \colon V_i \hookrightarrow V_j$ compatible with the local models, that is

$$\varphi_j \circ \psi_{ij} = \varphi_i.$$

Two such atlases are equivalent if the are contained in some bigger atlas. The space O equipped with an orbifold atlas is an *orbifold* of dimension n. See an example in Figure 3.18.

Remark 3.6.2. We think at the maps ψ_{ij} and f_{ij} as defined only up to the action of Γ_j (which acts on ψ_{ij} by composition and on f_{ij} by conjugation). In particular, if $U_i \subset U_j \subset U_k$ then we can verify that the equalities $\psi_{ik} = \psi_{jk} \circ \psi_{ij}$ and $f_{ik} = f_{jk} \circ f_{ij}$ hold only up to this ambiguity. See Figure 3.18.

The *isotropy group* Γ_p of a point $p \in O$ is the stabiliser of any lift of p in any local model V with respect to the action of Γ. By definition Γ_p is a finite subgroup of $O(n)$. A point p is *regular* if its isotropy group is trivial, and *singular* otherwise.

Example 3.6.3. The isotropy group O_p in the orbifold O from Figure 3.18 is $\mathbb{Z}_2 \times \mathbb{Z}_2$ at the origin, $\mathbb{Z}_2$ at the axis, and trivial elsewhere.

Proposition 3.6.4. *The regular points form a dense subset in O.*
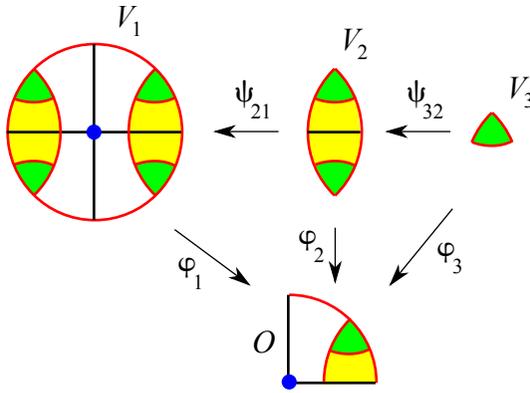
Figure 3.18. An orbifold O with three local models $O = U_1 \supset U_2 \supset U_3$. The group $\Gamma_1 = \mathbb{Z}_2 \times \mathbb{Z}_2$ acts on V_1 by reflections on the two coordinate axis, the group $\Gamma_2 = \mathbb{Z}_2$ acts on V_2 by reflection on the horizontal axis, Γ_3 is trivial.

Proof. On a local model $V \to V/\Gamma$, a singular point is the image of the fixed point locus of some element in Γ, which is in turn a proper subspace. $\square$

An orbifold is *locally oriented* if all the finite groups Γ lie in $\mathrm{SO}(n)$, and it is *oriented* if, in addition, all the transition maps ψ_{ij} are orientation-preserving. In a locally oriented orbifold reflections are not admitted and hence the singular locus has codimension $\geqslant 2$. An open subset of an orbifold is naturally an orbifold.

3.6.2. Examples. Orbifolds are natural objects and there are plenty of nice examples around.

Example 3.6.5. A differentiable manifold is an orbifold whose points are all regular. A differentiable manifold with boundary may be interpreted as an orbifold whose boundary points have the local structure of type $\mathbb{R}^n/\Gamma$ where $\Gamma = \mathbb{Z}_2$ is generated by a reflection along a hyperplane. The boundary should be interpreted as a *mirror*.

Exercise 3.6.6. Construct an orbifold structure on the triangle as suggested by Figure 3.19.

Example 3.6.7. Let $\Gamma < \mathrm{O}(n)$ be finite and $V \subset \mathbb{R}^n$ a Γ-invariant open set. The quotient V/Γ has an orbifold structure, defined by the unique local model $V \to V/\Gamma$.

We generalise the last example.

Proposition 3.6.8. *If M a Riemannian manifold and $\Gamma < \mathrm{Isom}(M)$ is a discrete subgroup, the quotient M/Γ has a natural orbifold structure.*
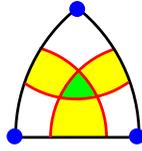
Figure 3.19. An orbifold structure on the triangle constructed using the local models from Figure 3.18. The isotropy group is $\mathbb{Z}_2 \times \mathbb{Z}_2$ at the vertices, $\mathbb{Z}_2$ at the edges, and trivial in the interior.

Proof. Take a point $p \in M/_\Gamma$ and $\tilde{p} \in M$ a lift. Since Γ is discrete, the stabiliser $\Gamma_{\tilde{p}}$ of $\tilde{p}$ is finite and there is a $r > 0$ such that $\exp_{\tilde{p}}(B_r(0)) = B_r(\tilde{p})$ and $g(B_r(\tilde{p}))$ intersects $B_r(\tilde{p})$ for some $g \in \Gamma$ if and only if $g \in \Gamma_{\tilde{p}}$.

The ball $B_r(\tilde{p})$ is clearly $\Gamma_{\tilde{p}}$-invariant. The group $\Gamma_{\tilde{p}}$ acts linearly and orthogonally on $B_r(0)$ and we get an orbifold local model

$$B_r(0) \longrightarrow B_r(0)/_{\Gamma_{\tilde{p}}} \cong B_r(\tilde{p})/_{\Gamma_{\tilde{p}}} = U_p$$

where the diffeomorphism is induced by the exponential map. We have constructed an orbifold local model U_p around p. We extend the covering $\{U_p\}$ thus obtained by adding all the non-empty intersections. $\square$

This is the richest source of nice examples. The quotient map $M \to M/_\Gamma$ is a covering, in an appropriate sense that we now explain.

3.6.3. Coverings. Let $\Gamma' < \Gamma < O(n)$ be finite groups and $V \subset \mathbb{R}^n$ a Γ-invariant open set. The natural map

$$\varphi \colon V/_{\Gamma'} \longmapsto V/_\Gamma$$

between the two orbifolds is a *local covering*.

Definition 3.6.9. A continuous map $p \colon \tilde{O} \to O$ between orbifolds is a *covering* if every point $p \in O$ has a neighbourhood U with $p^{-1}(U) = \sqcup_{i \in I} U_i$ and every restriction $p|_{U_i} \colon U_i \to U$ is a local covering.

The following is the main source of examples.

Example 3.6.10. Let M be a Riemannian manifold and $\Gamma' < \Gamma < \text{Isom}(M)$ be discrete groups. The natural map $M/_{\Gamma'} \to M/_\Gamma$ is an orbifold covering.

Definition 3.6.11 (The good, the bad, and the very good). An orbifold is *good* if it is covered by a manifold, and it is *bad* otherwise. It is *very good* if it is finitely covered by a manifold.

3.6.4. Hyperbolic, flat, and elliptic orbifolds. We define a *hyperbolic, flat*, or *elliptic* orbifold to be an orbifold whose local models are isometric quotients of balls in $\mathbb{H}^n$, $\mathbb{R}^n$, or S^n, and whose transition functions are also isometries.

The quotient of a hyperbolic, flat, or spherical manifold by a discrete group of isometries is naturally a hyperbolic, flat, or elliptic orbifold.

Example 3.6.12. The triangle group $\Gamma(a, b, c)$ introduced in Section 3.4.5 is a discrete group of isometries of S^2, $\mathbb{R}^2$, or $\mathbb{H}^2$. It defines a triangular orbifold Δ, which is hyperbolic, flat, or elliptic, according to whether $\frac{1}{a} + \frac{1}{b} + \frac{1}{c}$ is smaller, equal, or bigger than 1. The isotropy groups are the dihedral D_{2a}, D_{2b}, and D_{2c} at the vertices, $\mathbb{Z}_2$ at the edges, and trivial elsewhere.

The index-two subgroup $\Gamma^{\mathrm{or}}(a, b, c) \triangleleft \Gamma(a, b, c)$ consisting of orientation-preserving isometries is sometimes called a *von Dyck group*. It defines an orientable orbifold O that double-covers Δ. The orbifold O is a topological sphere with three singular points having rotation isotopy groups $\mathbb{Z}/_{a\mathbb{Z}}, \mathbb{Z}/_{b\mathbb{Z}}, \mathbb{Z}/_{c\mathbb{Z}}$. See Figure 3.20.

Exercise 3.6.13. Show that this is a presentation for $\Gamma^{\mathrm{or}}(a, b, c)$:

$$\langle\, r, s, t \mid r^a, s^b, t^c, rst \,\rangle.$$

The triple r, s, t of generators is intrinsically determined in the group up to simultaneous conjugation or inversion. Different unordered triples (a, b, c) produce non-isomorphic Von Dyck groups $\Gamma^{\mathrm{or}}(a, b, c)$.

Hints. The group $\Gamma(a, b, c)$ preserves the tessellation T of $\mathbb{H}^2$, $\mathbb{R}^2$, or S^2 into triangles with angles $\frac{\pi}{a}, \frac{\pi}{b}, \frac{\pi}{c}$, so its finite-order elements are rotations along some vertices of T. Three rotations r, s, t that satisfy $rst = 1$ are of some very special kind. $\square$

Example 3.6.14. More generally, every hyperbolic (or flat, elliptic) Coxeter polyhedron is a hyperbolic (or flat, elliptic) orbifold.

Example 3.6.15. Quotient the flat torus $T = \mathbb{R}^2/_{\mathbb{Z}^2}$ by the *elliptic involution* $(x, y) \mapsto (-x, -y)$. This isometry of T has four fixed points and the quotient flat orbifold is a sphere with four singular points, see Figure 3.21.

There are also many interesting non-compact orbifolds.

Example 3.6.16. A fundamental domain for $\mathbb{PSL}_2(\mathbb{Z})$ was shown in Figure 3.15. By gluing its sides according to the action we see quite easily that the hyperbolic orbifold $H^2/_{\mathbb{PSL}_2(\mathbb{Z})}$ is non-compact and contains two rotational points of order 2 and 3, see Figure 3.22.

Exercise 3.6.17. The 1-dimensional compact orbifolds are S^1 and the segment $[0, 1]$ with mirrored endpoints. Show that there are coverings $S^1 \to [0, 1]$ of degree 2, and also coverings $[0, 1] \to [0, 1]$ of any positive degree.
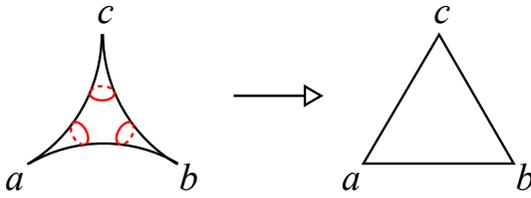
Figure 3.20. The index-two orientation-preserving $\Gamma^{\text{or}}(a, b, c) \lhd \Gamma(a, b, c)$ defines an index-two orbifold covering $O \to \Delta$, where O is a 2-sphere with three singular points and Δ a triangle.
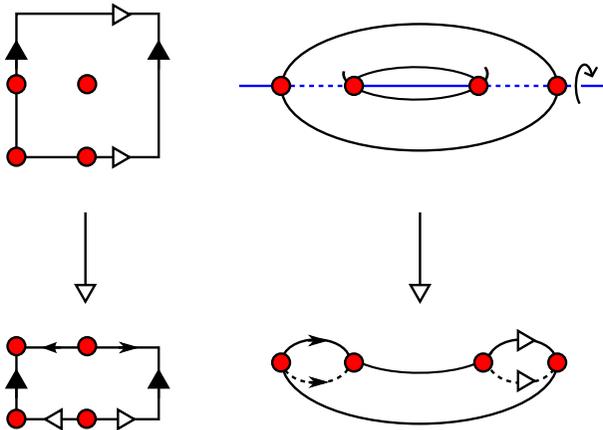
Figure 3.21. The elliptic involution quotients the torus to a sphere with four singular points. The fixed points and their images are drawn in red. We show both a planar (left) and spacial (right) picture.
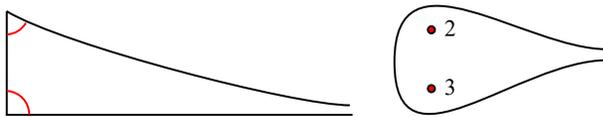
Figure 3.22. The hyperbolic orbifold $\mathbb{H}^2/_{\mathbb{PSL}_2(\mathbb{Z})}$ is obtained by mirroring the sides of a triangle with inner angles $\frac{\pi}{3}$, $\frac{\pi}{2}$, and zero (left). The orbifold is topologically a punctured sphere with two singular points with rotational isotropy $\mathbb{Z}_2$ and $\mathbb{Z}_3$ (right).

When $O = X/_\Gamma$ and X is a simply connected manifold, we say that Γ is the *fundamental group* of O. It is also possible to define the fundamental group for a generic orbifold using the appropriate orbi-notions of paths and homotopies.

3.6.5. Cone manifolds. Hyperbolic (flat, elliptic) cone manifolds and orbifolds are different objects, but they have a wide common intersection.

A hyperbolic (flat, elliptic) orbifold O whose isotropy groups O_p are either trivial or generated by a $\frac{2\pi}{p}$-rotation along a codimension-two subspace is naturally a hyperbolic (flat, elliptic) cone manifold with cone angles $\frac{2\pi}{p}$. The singular set may consist of various connected components and the natural number p depends on the component.

Conversely, every hyperbolic (flat, elliptic) manifold with cone angles that divide 2π can be given a natural hyperbolic (flat, elliptic) orbifold structure whose singular set consists of rotational points only. Moreover, we can easily prove that it is a good one:

Proposition 3.6.18. *Every hyperbolic (flat, elliptic) manifold M with cone angles that divide 2π is a good hyperbolic (flat, elliptic) orbifold. If M is complete, then $M = \mathbb{H}^n/_\Gamma$ ($\mathbb{R}^n/_\Gamma$, $S^n/_\Gamma$) for some discrete group Γ of isometries.*

Proof. We consider for simplicity only the hyperbolic case. We must prove that M is orbifold-covered by a hyperbolic manifold.

The singular set $S \subset M$ decomposes into connected components $S = \sqcup_i S_i$, each with a cone angle $\frac{2\pi}{p_i}$ for some integer $p_i \geqslant 2$. The complement $M' = M \setminus S$ is a non-complete hyperbolic manifold and as such it is equipped with a developing map $D \colon \tilde{M}' \to \mathbb{H}^n$ and a holonomy $\rho \colon \pi_1(M') \to \mathsf{Isom}(\mathbb{H}^n)$.

We pick a loop μ_i around each component S_i of S and note that by definition $\rho(\mu_i)$ is an elliptic isometry that rotates $\mathbb{H}^n$ around a codimension-two subspace by the angle $\frac{2\pi}{p_i}$. In particular $\rho(\mu_i)$ has order p_i.

We consider the regular covering $N \to M'$ corresponding to the subgroup $\ker \rho$ and give N the hyperbolic structure induced by M'. The completion $\overline{N}$ of N is a hyperbolic manifold that orbifold-covers M: since $\rho(\mu_i)$ has order precisely p_i, the added singular points in $\overline{N}$ have cone angle 2π and are hence ordinary.

If M is complete, then $\overline{N}$ also is and hence $\overline{N} = \mathbb{H}^n/_{\Gamma'}$. We define $\Gamma < \mathsf{Isom}(\mathbb{H}^n)$ to be the group generated by Γ' and any lifts of the deck transformations of $\overline{N} \to M$, and we get $M = \mathbb{H}^n/_\Gamma$. $\qquad\square$

Corollary 3.6.19. *If M is complete and has finitely generated fundamental group, then it is very good.*

Proof. We know that $M = \mathbb{H}^n/_\Gamma$ and we apply Selberg's Lemma. $\quad\square$

Summing up, complete hyperbolic cone manifolds with angles that divide 2π can be promoted to good orbifolds and hence to quotients $\mathbb{H}^n/_\Gamma$. We will encounter many examples in dimensions two and three later on.

3.6.6. References. All the material presented here is standard and introduced in various books, like Benedetti − Petronio [**4**] and Ratcliffe [**49**]. A beautiful introduction to regular polytopes is Coxeter's 1963 book [**14**], while the reader interested in Coxeter polytopes may consult Vinberg's survey [**59**].

The theory of orbifolds and cone manifolds presented here is quite limited: to get more background, the reader may consult Thurston's notes [**56**, Chapter 13] for the orbifolds, and McMullen's paper [**40**] for the cone manifolds.

Thick-thin decomposition

A peculiar aspect of hyperbolic geometry is the existence of complete hyperbolic manifolds that have finite volume but are not compact. These manifolds behave like the compact ones in many aspects, but are sometimes easier to construct. They arise naturally when we study hyperbolic surfaces, and are a fundamental constituent of Thurston's geometrisation of three-manifolds.

We prove here a structure theorem for all such manifolds. The theorem says that every finite-volume complete hyperbolic manifold can be decomposed into two domains: a *thick part* which is compact and has injectivity radius bounded from below, and a *thin part* that consists of *cusps*. A cusp is a (truncated) hyperbolic manifold of type $N \times [0, +\infty)$, where every section $N \times t$ has a flat metric that shrinks exponentially with t.

The core of this theorem is a lemma about discrete subgroups of Lie groups called the *Margulis Lemma*. We will apply this lemma also to the other geometries and prove Bierbach's Theorem, that states that every compact flat manifold is covered by a torus.

Throughout the discussion we will also study some general aspects of finite-volume complete hyperbolic manifolds, concerning in particular closed geodesics, isometry groups, and finite covers.

4.1. Tubes and cusps

We introduce here two very simple kinds of hyperbolic manifolds, so simple that their fundamental groups will be called *elementary* in the sequel: the tubes and the cusps. Before introducing them we show that the injectivity radius of a hyperbolic manifold behaves nicely.

We also study the closed geodesics in a hyperbolic manifold M, and prove that there is precisely one in every free homotopy class of closed curves of hyperbolic type.

4.1.1. Injectivity radius. Like many other geometric properties, the injectivity radius of a complete hyperbolic manifold may be nicely observed by looking at its universal cover. If $S \subset \mathbb{H}^n$ is a discrete set, we let $d(S)$ be the infimum of $d(x_1, x_2)$ among all pairs x_1, x_2 of distinct points in S.

Recall that every complete hyperbolic manifold is a quotient $M = \mathbb{H}^n/_\Gamma$ by some group Γ of isometries acting freely on $\mathbb{H}^n$.

Proposition 4.1.1. *Let $M = \mathbb{H}^n/_\Gamma$ be a complete hyperbolic manifold and $\pi\colon \mathbb{H}^n \to M$ the projection. For every $x \in M$ we have*

$$\text{inj}_x M = \frac{1}{2} \cdot d(\pi^{-1}(x)).$$

Proof. The number $\text{inj}_x M$ is the supremum of all $r > 0$ such that $B(x, r)$ is isometric to a ball of radius r in $\mathbb{H}^n$. The open set $B(x, r)$ is a ball of radius r if and only if its counterimage via π consists of disjoint balls of radius r, and this holds $\Leftrightarrow$ two distinct points in $\pi^{-1}(x)$ stay at distance at least $2r$. $\qquad\square$

Recall that $d(\gamma)$ is the minimum displacement of $\gamma \in \text{Isom}(\mathbb{H}^n)$.

Corollary 4.1.2. *If $M = \mathbb{H}^n/_\Gamma$ be a complete hyperbolic manifold, then*

$$\text{inj} M = \frac{1}{2} \cdot \inf \left\{ d(\gamma) \mid \gamma \in \Gamma, \gamma \neq \text{id} \right\}.$$

Proof. We have

$$\text{inj}_x M = \frac{1}{2} \cdot d(\pi^{-1}(x)) = \frac{1}{2} \cdot \inf \left\{ d(\tilde{x}, \gamma(\tilde{x})) \mid \gamma \in \Gamma, \gamma \neq \text{id}, \tilde{x} \in \pi^{-1}(x) \right\}.$$

Therefore

$$\text{inj} M = \inf_{x \in M} \text{inj}_x M = \frac{1}{2} \cdot \inf \left\{ d(\gamma) \mid \gamma \in \Gamma, \gamma \neq \text{id} \right\}.$$

The proof is complete. $\qquad\square$

Corollary 4.1.3. *If $M = \mathbb{H}^n/_\Gamma$ is a compact hyperbolic manifold then every non-trivial element in Γ is hyperbolic.*

Proof. Every nontrivial element in Γ is either hyperbolic or parabolic. If M is compact then $\text{inj} M > 0$. If Γ contains a parabolic γ then $d(\gamma) = 0$ and hence $\text{inj} M = 0$. $\qquad\square$

4.1.2. Tubes. Consider the infinite cyclic group $\Gamma = \langle \varphi \rangle$ generated by a hyperbolic transformation φ on $\mathbb{H}^n$ with axis l and minimum displacement $d > 0$. The iterates φ^k are again hyperbolic transformations with axis l and displacement kd. Therefore Γ acts freely on $\mathbb{H}^n$. The quotient manifold $M = \mathbb{H}^n/_\Gamma$ is called an *infinite tube*.

Exercise 4.1.4. Fix $q \in l$. Let q_1, q_2 be the two points in l at distance $\frac{d}{2}$ from q and π_1, π_2 the two hyperplanes orthogonal to l in q_1, q_2. The Dirichlet domain $D(q)$ of Γ is the space comprised between π_1 and π_2.

The infinite tube $M = \mathbb{H}^n/_\Gamma$ is obtained from $D(q)$ by identifying π_1 and π_2 along φ. Its fundamental group is isomorphic to $\Gamma \cong \mathbb{Z}$. The axis l projects in M onto a closed geodesic γ of length d. We have $\text{inj} M = \frac{d}{2}$ by Corollary 4.1.2 and the points in γ are precisely those with minimum injectivity radius. The closed geodesic γ is sometimes called the *core geodesic* of the tube.
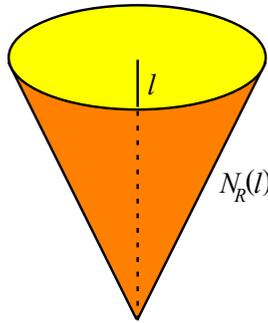
Figure 4.1. The R-neighbourhood $N_R(l)$ of a vertical line in the half-space model is a Euclidean cone. The boundary of the cone is not totally geodesic: the cone is however convex.

Proposition 4.1.5. *Every infinite tube is diffeomorphic to $S^1 \times \mathbb{R}^{n-1}$ or $S^1 \, \widetilde{\times} \, \mathbb{R}^{n-1}$ according to whether φ is orientation-preserving or not.*

Proof. By projecting $\mathbb{H}^n$ orthogonally onto l, we give $\mathbb{H}^n$ the structure of a $\mathbb{R}^{n-1}$-bundle over l which is preserved by φ and hence descends to a structure of $\mathbb{R}^{n-1}$-bundle over γ. The conclusion follows from the classification of vector bundles over S^1, see Proposition 1.1.13. $\qquad\square$

A *tube of radius R*, or a *R-tube*, is the quotient $N_R(l)/_\Gamma$ of the closed R-neighbourhood $N_R(l)$ of l, shown in Figure 4.1. It is diffeomorphic to $D^{n-1} \times S^1$ or $D^{n-1} \, \widetilde{\times} \, S^1$, and in particular it is compact. Note that the boundary of a tube is not totally geodesic, see Figure 4.1.

4.1.3. Cusps. We now introduce another simple type of hyperbolic manifolds. In the previous example the discrete group Γ consisted of hyperbolic transformations fixing the same line l, now Γ will consist of parabolic transformations fixing the same point at infinity.

Let $\Gamma < \mathrm{Isom}(\mathbb{R}^{n-1})$ be a non-trivial discrete group of Euclidean isometries acting freely on $\mathbb{R}^{n-1}$: the quotient $M = \mathbb{R}^{n-1}/_\Gamma$ is a flat $(n-1)$-manifold. If we use the half-space model for $\mathbb{H}^n$ with coordinates (x, t), every element $\varphi \in \Gamma$ acts as a parabolic transformation on $\mathbb{H}^n$ by sending (x, t) to $(\varphi(x), t)$. The whole group Γ is a discrete group of parabolic transformations of $\mathbb{H}^n$ fixing the point ∞.

The quotient $\mathbb{H}^n/_\Gamma$ is naturally diffeomorphic to $M \times \mathbb{R}_{>0}$. The metric tensor at the point (x, t) is

$$g_{(x,t)} = \frac{g_x^M \oplus 1}{t^2}$$

where g^M is the metric tensor of the flat M. The manifold $\mathbb{H}^n/_\Gamma$ is called a *cusp*. Since Γ contains parabolics we have $\mathrm{inj} M = 0$.

Figure 4.2. The *pseudosphere* is a surface in $\mathbb{R}^3$ isometric
to the union of two truncated cusps, each with constant
gaussian curvature -1.

Remark 4.1.6. The vertical coordinate t may be parametrized more
intrinsically using arc-length. As we have seen in Proposition 2.1.24, a
vertical geodesic with unit speed is parametrized as $t = e^u$. Using u
instead of t the cusp is isometric to $M \times \mathbb{R}$ with metric tensor

$$g_{(x,u)} = (e^{-2u} g_x^M) \oplus 1.$$

The lengths in the flat slice $M \times u$ are shrunk or dilated by the factor e^{-u}.

A *truncated cusp* is a portion $N = M \times [a, +\infty)$, bounded by the
Euclidean manifold $M \times a$: note that the boundary ∂N is Euclidean but
not totally geodesic. The volume of a truncated cusp is strikingly simple.

Proposition 4.1.7. *Let N be a truncated cusp. We have*

$$\mathrm{Vol}(N) = \frac{\mathrm{Vol}(\partial N)}{n - 1}.$$

Proof. It follows from Lemma 3.2.7.   □

In particular a (non truncated) cusp has infinite volume.

Example 4.1.8. In dimension $n = 2$ there is only one cusp up to
isometry. The group $\Gamma < \mathrm{Isom}(\mathbb{R})$ is the infinite cyclic group generated by
a translation $x \mapsto x + b$ and up to conjugating in $\mathrm{Isom}(\mathbb{H}^2)$ we may take
$b = 1$. The cusp is diffeomorphic to $S^1 \times \mathbb{R}$, and the circle $S^1 \times \{u\}$ has
length e^{-u}. Some truncated cusp (but not the whole cusp!) embeds in $\mathbb{R}^3$
as in Figure 4.2.

Remark 4.1.9. Pick $p \in \mathbb{H}^2$. Note that a cusp and $\mathbb{H}^2 \setminus \{p\}$ are both
hyperbolic and diffeomorphic to an open annulus $S^1 \times \mathbb{R}$. However, they
are not isometric because the cusp is complete while $\mathbb{H}^2 \setminus \{p\}$ is not.

It is customary to employ the word *cusp* to indicate a truncated cusp.

4.1.4. Closed geodesics. A *closed curve* in a manifold M is a smooth map $\gamma \colon S^1 \to M$. A (possibly closed) curve is *simple* if it is an embedding.

We consider S^1 as the unit circle in $\mathbb{C}$. A *closed geodesic* in a Riemannian manifold M is a smooth map $\gamma \colon S^1 \to M$ whose lift $\gamma \circ \pi \colon \mathbb{R} \to M$ along the universal covering $\pi(t) = e^{it}$ is a non-constant geodesic. Two closed geodesics γ_1, γ_2 that differ only by a rotation, *i.e.* such that $\gamma_1(z) = \gamma_2(ze^{it})$ for some fixed $t \in \mathbb{R}$, are implicitly considered equivalent. The closed geodesics $\gamma(z)$ and $\overline{\gamma}(z) = \gamma(\bar{z})$ are however distinct (they have opposite orientations).

Proposition 4.1.10. *Let γ be a closed geodesic in a Riemannian manifold M. Exactly one of the following holds:*

  (1) *the curve γ is simple,*
  (2) *the curve γ self-intersects transversely in finitely many points,*
  (3) *the curve γ wraps $k \geqslant 2$ times along a curve of type (1) or (2).*

Proof. If the geodesic is not simple, it self-intersects. If it self-intersects only with distinct tangents, then (2) holds. Otherwise (3) holds. $\square$

The natural number k in (3) is the *multiplicity* of the closed geodesic. A closed geodesic γ of multiplicity k is of type $\gamma(e^{it}) = \eta(e^{kit})$ for some geodesic η of type (1) or (2).

Remark 4.1.11. A closed geodesic on a Riemannian manifold M is determined by its support, its orientation, and its multiplicity.

4.1.5. Closed geodesics in a hyperbolic manifold. Closed geodesics in hyperbolic manifolds have a particularly nice behaviour.

Let X, Y be topological spaces: as usual we indicate by $[X, Y]$ the homotopy classes of continuous maps $X \to Y$. Let X be path-connected. There is a natural map $\pi_1(X, x_0) \to [S^1, X]$, and the following is a standard exercise in topology.

Exercise 4.1.12. The map induces a bijection between the conjugacy classes in $\pi_1(X, x_0)$ and $[S^1, X]$.

A simple closed curve in X is *homotopically trivial* if it is homotopic to a constant. As a corollary, a simple closed curve γ is homotopically trivial if and only if it represents the trivial element in $\pi_1(X, \gamma(1))$.

On a complete hyperbolic manifold $M = \mathbb{H}^n / \Gamma$ the conjugacy classes of $\pi_1(M)$ correspond to those in Γ and we get a correspondence

$$\{\text{conjugacy classes in } \Gamma\} \longleftrightarrow [S^1, M].$$

This correspondence works as follows: given $\varphi \in \Gamma$, pick any $x \in \mathbb{H}^n$, connect x to $\varphi(x)$ with any arc, and project it to a closed curve in M.

Two conjugate elements in Γ are of the same type (trivial, parabolic, or hyperbolic) and have the same minimum displacement. Therefore every element in $[S^1, M]$ has a well-defined type and minimum displacement.

Proposition 4.1.13. *Let M be a complete hyperbolic manifold. Every hyperbolic element of $[S^1, M]$ is represented by a unique closed geodesic, of length d equal to its minimum displacement. The trivial and parabolic elements are not represented by closed geodesics.*

Proof. Take $M = \mathbb{H}^n/_\Gamma$. A hyperbolic isometry $\varphi \in \Gamma$ has a unique invariant geodesic in $\mathbb{H}^n$, its axis, which projects to a closed geodesic of length d. Conjugate isometries determine the same closed geodesic in M.

On the other hand, a closed geodesic in M lifts to a segment connecting two distinct points x_0 and $\varphi(x_0)$ for some $\varphi \in \Gamma$ which preserves the line passing through x_0 and $\varphi(x_0)$: since φ preserves a line, it is hyperbolic. $\square$

We get a bijection

$$\{\text{hyperbolic conjugacy classes in } \Gamma\} \longleftrightarrow \{\text{closed geodesics in } M\}.$$

Corollary 4.1.14. *Let M be a closed hyperbolic manifold. Every non-trivial element in $[S^1, M]$ is represented by a unique closed geodesic.*

Proof. Since M is compact there are no parabolics. $\square$

Corollary 4.1.15. *Let M be a complete hyperbolic manifold. Every closed geodesic has the minimum length among the closed curves in its homotopy class.*

Proof. If α is a closed geodesic, its length equals the minimum displacement d of a corresponding hyperbolic transformation φ. Every closed curve β homotopic to α lifts to an arc connecting two points $\tilde{x}$ and $\varphi(\tilde{x})$ that have distance at least d; hence β has length at least d. $\square$

Not only a purely topological object like a hyperbolic homotopy class of closed curves has a unique geometric nice representative, but this representative is the shortest possible one. Informally, we may think that every closed curve may be shrunk until it becomes a closed geodesic, and the negative curvature forces this closed geodesic to be unique: we will soon see that the uniqueness is lost in the elliptic and flat geometries.

It is worth recalling that a closed geodesic may not be simple. When it is simple, we now show that the closed geodesic has some nice small neighbourhoods. We defined the R-tubes in Section 4.1.2.

Proposition 4.1.16. *The R-neighbourhood of a simple closed geodesic γ in a complete hyperbolic manifold is isometric to a R-tube, if $R > 0$ is sufficiently small.*

Proof. By compactness of γ there is a sufficiently small $R > 0$ such that the R-neighbourhood of γ lifts to disjoint R-neighbourhoods of its geodesic lifts in $\mathbb{H}^n$. Hence their quotient is a R-tube. $\square$

4.2. The Margulis Lemma

We state and prove the *Margulis Lemma*, that concerns arbitrary discrete groups in Lie groups and more specifically in $\mathrm{Isom}(\mathbb{H}^n)$. The lemma implies that there is a constant $\varepsilon > 0$, which depends only on the dimension n, such that the ε-thin part of any complete hyperbolic n-manifold consists of truncated cusps and tubes only. The ε-thin part is by definition the set of all the points with injectivity radius smaller than ε.

The proof of the lemma for general Lie groups is surprisingly simple and elegant; its application to the hyperbolic case needs however a more technical argument. We start by exposing some preliminary facts on the isometries of the hyperbolic space that are of independent interest.

4.2.1. Isometries that commute or generate discrete groups. Two
isometries of $\mathbb{H}^n$ that commute or generate a discrete group must be of a particular kind. Let $\mathrm{Fix}(\varphi)$ denote the fixed points in $\overline{\mathbb{H}^n}$ of an isometry φ.

Lemma 4.2.1. *Let $\varphi_1, \varphi_2 \in \mathrm{Isom}(\mathbb{H}^n)$ be two hyperbolic or parabolic isometries. If they commute then $\mathrm{Fix}(\varphi_1) = \mathrm{Fix}(\varphi_2)$.*

Proof. If they commute, the map φ_1 acts on $\mathrm{Fix}(\varphi_2)$ and vice versa. If φ_2 is hyperbolic, then $\mathrm{Fix}(\varphi_2) = \{p, q\}$ and φ_1 fixes the line with endpoints p and q, hence is again hyperbolic with $\mathrm{Fix}(\varphi_1) = \{p, q\}$. If φ_1 and φ_2 are parabolic then they have the same fixed point $\mathrm{Fix}(\varphi_1) = \mathrm{Fix}(\varphi_2)$. $\square$

Lemma 4.2.2. *Let $\varphi_1, \varphi_2 \in \Gamma$ be two non-trivial isometries in a discrete group $\Gamma < \mathrm{Isom}(\mathbb{H}^n)$ that acts freely on $\mathbb{H}^n$. Either $\mathrm{Fix}(\varphi_1) \cap \mathrm{Fix}(\varphi_2) = \varnothing$, or one of the following holds:*

- *φ_1 and φ_2 are parabolics with the same fixed point,*
- *φ_1 and φ_2 are powers of the same hyperbolic $\varphi \in \Gamma$.*

Proof. Suppose that φ_1 and φ_2 have some common fixed points. We first consider the case where φ_1 is hyperbolic and φ_2 is parabolic. We pick the half-space model and suppose $\mathrm{Fix}(\varphi_1) = \{0, \infty\}$ and $\mathrm{Fix}(\varphi_2) = \{\infty\}$. Proposition 2.2.8 says that

$$\varphi_1(x, t) = \lambda(Ax, t), \quad \varphi_2(x, t) = (A'x + b, t)$$

with $A, A' \in O(n - 1)$ and $\lambda \neq 1$. Hence

$$\varphi_1^n \circ \varphi_2 \circ \varphi_1^{-n}(x, t) = \varphi_1^n\big(A'(\lambda^{-n}A^{-n}x) + b, \lambda^{-n}t\big)$$
$$= \big(A^n A' A^{-n}x + \lambda^n A^n b, t\big).$$

Up to interchanging φ_1 and φ_1^{-1} we may suppose $\lambda < 1$ and get

$$\lim_{n \to \infty} \varphi_1^n \circ \varphi_2 \circ \varphi_1^{-n}(0, t) = \lim_{n \to \infty} (\lambda^n A^n b, t) = (0, t).$$

A contradiction since Γ is discrete.

We now suppose that both φ_1 and φ_2 are hyperbolic, with $\mathrm{Fix}(\varphi_1) = \{a, \infty\}$ and $\mathrm{Fix}(\varphi_2) = \{b, \infty\}$. The isometries φ_1 and φ_2 permute the horospheres centred at ∞ and

$$[\varphi_1, \varphi_2] = \varphi_1 \circ \varphi_2 \circ \varphi_1^{-1} \circ \varphi_2^{-1} \in \Gamma$$

fixes every horizontal horosphere. Hence the commutator is parabolic or trivial: the first case is excluded by the previous discussion, in the second case we have $a = b$ by Lemma 4.2.1.

Both φ_1 and φ_2 have the same axis l, and since they generate a discrete group $\Gamma_0 < \Gamma$ they are both powers of some hyperbolic $\varphi \in \Gamma_0$ with that axis. To prove this, note that Γ_0 acts effectively on l as a discrete group of translations, hence $\Gamma_0 \cong \mathbb{Z}$. $\qquad\square$

Two isometries are contained in some discrete group if and only if they generate a discrete group, so the previous lemma is actually a fact on pairs of isometries that generate discrete groups. The two lemmas have important geometric consequences.

Corollary 4.2.3. *Let $\mathbb{H}^n/_\Gamma$ be a complete hyperbolic manifold. The axis in $\mathbb{H}^n$ of two hyperbolic isometries in Γ are either incident or ultra-parallel (not asymptotically parallel).*

Corollary 4.2.4. *Let $\mathbb{H}^n/_\Gamma$ be a complete hyperbolic manifold. Every subgroup of Γ isomorphic to $\mathbb{Z} \times \mathbb{Z}$ consists of parabolic elements fixing the same point at infinity.*

Corollary 4.2.5. *The fundamental group $\pi_1(M)$ of a closed hyperbolic manifold M does not contain subgroups isomorphic to $\mathbb{Z} \times \mathbb{Z}$.*

Corollary 4.2.6. *The torus has no hyperbolic structure.*

4.2.2. The Margulis Lemma. We now introduce the main character of this chapter, the Margulis Lemma. We first state a general version for Lie groups and then turn more specifically to $\mathrm{Isom}(\mathbb{H}^n)$.

Lemma 4.2.7. *Let G be a Lie group. There is a neighbourhood U of $e \in G$ such that every discrete subgroup $\Gamma < G$ generated by some elements in U is nilpotent.*

Note that the discreteness of Γ is essential here, since every connected Lie group G is generated by arbitrarily small elements, and G needs not to be nilpotent.

Proof. Consider the commutator map

$$[\ ,\ ]\colon G \times G \to G$$

that sends (g, h) to the commutator $[g, h]$. The map is smooth and sends $G \times \{e\}$ and $\{e\} \times G$ to the point e. Therefore its differential at (e, e)

vanishes. We identify a neighbourhood of e in G with $\mathbb{R}^n$, so that the commutator map near (e, e) can be read as

$$[\,,\,]\colon U \times U \longrightarrow \mathbb{R}^n$$

for some neighbourhood $U \subset \mathbb{R}^n$ of the origin. Since its differential at $(0, 0)$ vanishes, up to restricting U we may suppose that the map is $\frac{1}{2}$-Lipschitz. Therefore for all $x, y \in U$ we get

$$\|[x, y]\| < \|[x, 0]\| + \frac{1}{2}\|y\| = \frac{1}{2}\|y\|, \quad \|[x, y]\| < \|[0, y]\| + \frac{1}{2}\|x\| = \frac{1}{2}\|x\|$$

and hence

$$\|[x, y]\| < \frac{1}{2}\min\{\|x\|, \|y\|\}.$$

This implies that for every smaller neighbourhood $V \subset U$ of 0 there is a $k > 0$ such that

$$\underbrace{[U, [U, \ldots [U, U]]\cdots]}_{k} \subset V.$$

We now turn back to G. Let Γ be discrete and generated by some elements $S \subset U$. We choose a smaller neighbourhood V of e such that $V \cap \Gamma = \{e\}$, hence for any $a_1, \ldots, a_k, b \in S$ we get $[a_1, [a_2, \ldots [a_k, b]]\cdots] \in V \cap \Gamma = \{e\}$. Proposition 1.4.5 says that Γ is nilpotent. $\square$

We now want to refine this lemma when $G = \mathrm{Isom}(\mathbb{H}^n)$. Let P be a property of groups, like being abelian, nilpotent, etc. A group is *virtually P* if it has a finite-index subgroup which is P.

Lemma 4.2.8 (Margulis Lemma). *In every dimension $n \geqslant 2$ there is a constant $\varepsilon_n > 0$ such that for all $x \in \mathbb{H}^n$, every discrete group $\Gamma < \mathrm{Isom}(\mathbb{H}^n)$ generated by elements that move x at distance smaller than ε_n is virtually nilpotent.*

Proof. It suffices to prove the theorem for a fixed $x \in \mathbb{H}^n$ since the isometries of $\mathbb{H}^n$ act transitively on points. By Lemma 4.2.7 there is a neighbourhood U of e in $G = \mathrm{Isom}(\mathbb{H}^n)$ such that every discrete group generated by some elements in U is nilpotent.

Let $V_\varepsilon \subset \mathrm{Isom}(\mathbb{H}^n)$ be the set of all the isometries that move x at distance smaller than $\varepsilon > 0$. As $\varepsilon > 0$ varies, the relatively compact sets V_ε form a neighbourhood system for the compact stabiliser $G_x < \mathrm{Isom}(\mathbb{H}^n)$ of x, isomorphic to $\mathrm{O}(n)$. We can prove (exercise) that $V_\varepsilon^{-1} = V_\varepsilon$ and

$$V_\varepsilon^k = V_{k\varepsilon}$$

for every positive $k \in \mathbb{N}$. Since G_x is compact, it is covered by some m translates of U. Pick a $\varepsilon > 0$ such that V_ε is contained in the union of these m translates. We prove the theorem with $\varepsilon_n = \varepsilon/m$.

Let Γ be a discrete group generated by some elements $S \subset V_{\varepsilon_n}$. We need to show that Γ is virtually nilpotent. In fact we prove that the nilpotent subgroup Γ_U generated by $\Gamma \cap U$ has index at most m in Γ.

Suppose by contradiction that $\Gamma/_{\Gamma_U}$ contains more than m cosets. Let $r(i)$ be the number of cosets represented by elements in Γ that are products of at most i generators in S. We have $r(1) > 1$, otherwise $\Gamma = \Gamma_U$. If $r(i) = r(i+1)$ for some i, then it is easy to deduce that $r(i)$ stabilises forever. This easily implies that $r(m) > m$. Therefore there are $m+1$ elements in $V_{\varepsilon_n}^m = V_{m\varepsilon_n} = V_\varepsilon$ belonging to $m+1$ distinct cosets. By hypothesis two of them, say g and h, belong to the same translate of U covering V_ε. Therefore $gh^{-1} \in \Gamma \cap U$, contradicting the fact that g and h lie in distinct cosets. $\square$

For a discrete group $\Gamma < \mathrm{Isom}(\mathbb{H}^n)$ and a point $x \in \mathbb{H}^n$, we denote by $\Gamma_\varepsilon(x) < \Gamma$ the subgroup generated by all elements $g \in \Gamma$ that move x of a distance smaller than ε. We have proved that $\Gamma_{\varepsilon_n}(x)$ is virtually nilpotent for every $x \in \mathbb{H}^n$ and every discrete Γ.

A *Margulis constant* in a fixed dimension n is any number $\varepsilon_n > 0$ for which Lemma 4.2.8 holds. Every sufficiently small number is a Margulis constant.

4.2.3. Elementary groups. A non-trivial discrete group $\Gamma < \mathrm{Isom}(\mathbb{H}^n)$ is *elementary* if it preserves a finite set of points in $\overline{\mathbb{H}^n}$.

Proposition 4.2.9. *An elementary Γ acting freely on $\mathbb{H}^n$ is:*

- *generated by a hyperbolic isometry, or*
- *generated by parabolic isometries having the same fixed point at ∞.*

Proof. Since Γ contains no elliptics, every non-trivial element fixes one or two points at infinity in $\overline{\mathbb{H}^n}$, and no other finite set of points (because of Proposition 2.4.6). We conclude using Lemma 4.2.2. $\square$

Proposition 4.2.10. *Let $\Gamma < \mathrm{Isom}(\mathbb{H}^n)$ be a discrete group acting freely on $\mathbb{H}^n$. If $\Gamma' < \Gamma$ has finite index and is elementary, then Γ also is.*

Proof. We know that every element in Γ' is either hyperbolic with axis l, or parabolic with fixed point $p \in \partial\mathbb{H}^n$. If $\varphi \in \Gamma$ then $\varphi^k \in \Gamma'$ for some k: hence φ is also of that type. $\square$

Corollary 4.2.11. *Let $\Gamma < \mathrm{Isom}(\mathbb{H}^n)$ be discrete and acting freely. If Γ is virtually nilpotent, it is either trivial or elementary.*

Proof. We know that Γ contains a finite-index nilpotent subgroup H. If H is trivial then Γ is finite and hence trivial. If H is non-trivial, it has a non-trivial centre by Proposition 1.4.6. Lemma 4.2.1 then implies that all the elements in H have the same fixed points and so H is elementary. Therefore Γ is elementary by Proposition 4.2.10 $\square$

Let ε_n be a Margulis constant. We can strengthen the Margulis Lemma.

Corollary 4.2.12. *Let $\Gamma < \mathrm{Isom}(\mathbb{H}^n)$ be discrete and acting freely. For every point $x \in \mathbb{H}^n$ the subgroup $\Gamma_{\varepsilon_n}(x)$ is either trivial or elementary.*

4.2.4. Thick-thin decomposition. A notable geometric consequence of the Margulis Lemma is that every complete hyperbolic manifold decomposes into a thick and a thin part, where the thin part consists only of particularly simple star-shaped pieces. This decomposition is known as the *thick-thin decomposition* of hyperbolic manifolds.

We define a *star-shaped set* centred at $p \in \partial\mathbb{H}^n$ to be any subset $U \subset \mathbb{H}^n$ that intersects every half-line pointing to p in a half-line. For instance, a horoball is star-shaped. A *star-shaped neighbourhood* of a line $l \subset \mathbb{H}^n$ is any neighbourhood V of l that intersects every line orthogonal to l into a connected set. For instance, a R-neighbourhood of l is star-shaped.

These definitions pass to quotients. A *star-shaped cusp neighbourhood* is the quotient $U/_\Gamma$ of a Γ-invariant star-shaped set U centred at $p \in \partial\mathbb{H}^n$ via a discrete group Γ of parabolic transformations fixing p and acting freely. Analogously, a *star-shaped simple closed geodesic neighbourhood* is the quotient $V/_\Gamma$ of a Γ-invariant star-shaped neighbourhood V of l via a discrete group $\Gamma \cong \mathbb{Z}$ of hyperbolic transformations with axis l.

The truncated cusps and R-tubes studied in Section 4.1 are particularly nice star-shaped cusp and geodesic neighbourhoods.

Let $\varepsilon_n > 0$ be a Margulis constant. We define $M_{[\varepsilon_n,\infty)}$ and $M_{(0,\varepsilon_n]}$ respectively as the set of all points $x \in M$ having $\mathrm{inj}_x M \geqslant \frac{\varepsilon_n}{2}$, and as the closure of the complementary set $M \setminus M_{[\varepsilon_n,\infty)}$. They form respectively the *thick* and *thin part* of M.

Remark 4.2.13. We do not define the thin part simply as the set of all points x having $\mathrm{inj}_x M \leqslant \frac{\varepsilon_n}{2}$ because we want to discard the degenerate and unlucky case of a closed geodesic γ having length precisely ε_n. The injectivity radius would be $\frac{\varepsilon_n}{2}$ at the points in γ and strictly bigger than $\frac{\varepsilon_n}{2}$ near γ. With our definition, the geodesic γ is contained in $M_{[\varepsilon_n,+\infty)}$ and not in $M_{(0,\varepsilon_n]}$.

The following theorem is arguably the most important structural result on complete hyperbolic manifolds of any dimension n.

Theorem 4.2.14 (Thick-thin decomposition). *Let M be a complete hyperbolic n-manifold. The thin part $M_{(0,\varepsilon_n]}$ consists of a disjoint union of star-shaped neighbourhoods of cusps and of simple closed geodesics of length $< \varepsilon_n$.*

Proof. We have $M = \mathbb{H}^n/_\Gamma$. For every isometry $\varphi \in \Gamma$ we define

$$S_\varphi(\varepsilon) = \left\{ x \in \mathbb{H}^n \mid d(\varphi(x), x) \leqslant \varepsilon \right\} \subset \mathbb{H}^n.$$

By Proposition 4.1.1 the thin part $M_{(0,\varepsilon_n]}$ is the image of the set

$$S = \left\{ x \in \mathbb{H}^n \mid \exists \varphi \in \Gamma, \varphi \neq \mathrm{id} \text{ such that } d(\varphi(x), x) \leqslant \varepsilon_n \right\}$$

$$= \bigcup_{\varphi \in \Gamma, \varphi \neq \mathrm{id}} S_\varphi(\varepsilon_n).$$

More precisely, we should exclude the hyperbolic $\varphi \in \Gamma$ with $d(\varphi) = \varepsilon_n$, see Remark 4.2.13, but this is not an important point and we ignore it. It is easy to check that $S_\varphi(\varepsilon)$ is star-shaped, centred at a $p \in \partial\mathbb{H}^n$ or at a line l according to whether φ is parabolic fixing p or hyperbolic fixing l.

Suppose that $x \in S_\varphi(\varepsilon_n) \cap S_\psi(\varepsilon_n)$ for some non-trivial isometries $\varphi, \psi \in \Gamma$. By the Margulis Lemma both φ and ψ belong to the elementary group $\Gamma_{\varepsilon_n}(x)$ and hence by Proposition 4.2.9 both φ and ψ are either parabolic fixing the same point p at infinity or hyperbolic fixing the same line l.

Therefore every connected component S_0 of S is the union of all $S_\varphi(\varepsilon_n)$ where φ varies in some maximal elementary subgroup $\Gamma_0 < \Gamma$ of parabolics fixing the same point p or hyperbolics fixing the same line l. The set S_0 is a union of star-shaped sets centred at p or l and is hence also star-shaped.

The group Γ preserves S and the only isometries in Γ that preserve S_0 are those in Γ_0, therefore the quotient $M_{(0,\varepsilon_n]} = S/\Gamma$ consists of star-shaped neighbourhoods of cusps and of simple closed geodesics. $\qquad\square$

Star-shaped neighbourhoods are particularly nice in low dimensions.

Proposition 4.2.15. *Let M be a complete orientable hyperbolic manifold of dimension $n \leqslant 3$. The thin part $M_{(0,\varepsilon_n]}$ consists of truncated cusps and R-tubes.*

Proof. Pick a non-trivial $\varphi \in \mathrm{Isom}^+(\mathbb{H}^n)$. We check that $S_\varphi(\varepsilon)$ is either empty, or a R-neighbourhood of a line l, or a horoball, for all $\varepsilon > 0$. This proves the proposition.

Suppose that φ is hyperbolic with axis l. The distance $d(x, \varphi(x))$ depends only on $d(x, l)$ since all the orientation-preserving hyperbolic transformations with axis l commute with φ and act transitively on the points at fixed distance from l. It is easy to check that $d(x, \varphi(x))$ increases with $d(x, l)$ and hence $S_\varphi(\varepsilon)$ is either empty or a R-neighbourhood of l.

If φ is parabolic, it acts on each horosphere centred at p like a Euclidean fixed-point-free orientation-preserving isometry on $\mathbb{R}^{n-1}$: this must be a translation when $n \leqslant 3$. Therefore $d(x, \varphi(x))$ depends only on the horosphere O containing x and decreases as O moves towards p. Therefore $S_\varphi(\varepsilon)$ is a horoball. $\qquad\square$

See the picture in Figure 4.3.

Corollary 4.2.16. *Let M be a complete hyperbolic n-manifold. The closed geodesics in M of length $< \varepsilon_n$ are simple and disjoint.*
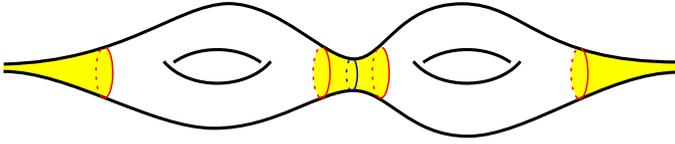
Figure 4.3. The thick-thin decomposition of a complete hyperbolic surface: the thin part (yellow) consists of truncated cusps and neighbourhoods of short geodesics (blue).

Proof. These closed geodesics lie in the thin part. Star-shaped cusp neighbourhoods contain no closed geodesics, and each star-shaped geodesic neighbourhood contains only one closed geodesic, its core, which is simple.   □

4.2.5. Finite-volume hyperbolic manifolds. A specific feature of hyperbolic geometry is the existence of complete hyperbolic manifolds that have finite volume without being compact. The thin-thick decomposition furnishes a nice topological description of such manifolds.

Proposition 4.2.17. *A complete hyperbolic manifold M has finite volume if and only if its thick part is compact.*

Proof. If the thick part is not compact, it contains an infinite number of points that stay pairwise at distance greater than ε_n. The open balls of radius $\frac{\varepsilon_n}{2}$ centred at these points are embedded and disjoint and all have the same volume: therefore their union has infinite volume.

If the thick part is compact, it has finite volume. Its boundary is also compact, and hence has finitely many connected components. Therefore the thin part consists of finitely many star-shaped neighbourhoods of cusps and closed geodesics, each with compact boundary. Each such object has finite volume (because it is contained in a bigger abstract truncated cusp with compact base, or a R-tube, which has finite volume).   □

Corollary 4.2.18. *Every complete finite-volume hyperbolic manifold M is diffeomorphic to the interior of a compact manifold N with boundary. The boundary ∂N consists of manifolds that admit some flat structure.*

Proof. We have seen in the previous proof that the thick part of M is compact, and the thin part consists of finitely many star-shaped neighbourhoods of cusps and closed geodesics, each with compact boundary.

Every cusp neighbourhood has compact base and hence contains a smaller truncated cusp, diffeomorphic to $X \times [0, 1)$ for some closed flat manifold X. The complement in M of these truncated cusps is compact: it is obtained from the thick part by adding finitely many neighbourhoods of closed geodesics (that are compact) and compact portions of cusp neighbourhoods. Each truncated cusp is diffeomorphic to $X \times [0, 1)$ and can be

compactified by adding $X \times 1$. The resulting manifold N is compact with boundary.   □

Every boundary component X of N inherits a flat structure, uniquely determined up to rescaling (different truncations modify X only by rescaling.) As already mentioned, a truncated cusp is often called simply a *cusp*: for instance we say that the surface sketched in Figure 4.3 has two cusps.

We briefly discuss the effects of Corollary 4.2.18 on low-dimensional manifolds. Let M be a complete hyperbolic manifold of finite volume of dimension n. If $n = 2$, the manifold M is diffeomorphic to the interior of a compact surface N with boundary. Every boundary component of N is of course a circle.

If $n = 3$, the manifold M is diffeomorphic to the interior of a compact 3-manifold N with boundary, and every boundary component X of N is a flat surface: if M is orientable, then ∂N also is, and by Proposition 3.4.12 it consists of tori.

4.2.6. Geodesic boundary and cusps. Most of the arguments of this chapter extend to hyperbolic manifolds with compact geodesic boundary.

Proposition 4.2.19. *Every complete finite-volume hyperbolic manifold M with compact geodesic boundary is diffeomorphic to a compact manifold N with some boundary components removed. The removed components have a flat structure. The remaining components form the geodesic boundary of M and hence have a hyperbolic structure.*

Proof. Double M along the geodesic boundary to obtain a finite-volume hyperbolic manifold to which Corollary 4.2.18 applies.   □

We briefly discuss the effects in low dimensions. In dimension $n = 2$ the boundary ∂N consists of circles: some are geodesic components of ∂M, while some others are removed and correspond to cusps. In dimension $n = 3$, if M is orientable the removed boundary of ∂N consists of tori (the cusps), while the geodesic boundary of M cannot contain any torus: a torus has no hyperbolic structure by Corollary 4.2.6.

We add some information on fundamental groups. Let the manifolds M and N be as in Proposition 4.2.19.

Proposition 4.2.20. *For every boundary component X of N the homomorphism $\pi_1(X) \to \pi_1(N)$ induced by inclusion is injective. Two homotopically non-trivial closed curves in distinct boundary components of N are not freely homotopic in N.*

Proof. We have $M = C/\Gamma$ for some convex $C \subset \mathbb{H}^3$ with boundary consisting of hyperplanes. If X is a (cusp) flat component of ∂N, it is isometric to $\mathbb{R}^{n-1}/\Gamma_X$ and the map $\Gamma_X = \pi_1(X) \to \pi_1(N) = \Gamma$ is injective since it sends a Euclidean isometry to a corresponding parabolic isometry.

If X is geodesic, it lifts to a hyperplane in ∂C, which is simply connected: hence $\pi_1(X) \to \pi_1(N)$ is injective.

A homotopically non-trivial closed curve $\gamma \subset X$ determines a conjugacy class of hyperbolic or parabolic transformations in Γ. Every element in this class preserves a unique boundary hyperplane or a point at infinity of C that projects back to X, so the conjugacy class determines X. Therefore distinct boundary components contain different conjugacy classes. $\square$

We note that a finite-volume hyperbolic manifold M may have non-compact geodesic boundary: an ideal polygon in $\mathbb{H}^2$ is a simple example.

4.3. Geodesic spectrum, isometry groups, and finite covers

We now study the geodesic spectrum and the isometry group of hyperbolic manifolds. The geodesic spectrum of M is the set of the lengths of all the closed geodesics in M. We prove that a finite-volume M has a discrete geodesic spectrum (with finite multiplicities) and a finite isometry group.

Then we turn to finite covers and derive some geometric consequences from the residually finiteness of fundamental groups.

We are almost exclusively interested in finite-volume hyperbolic manifolds. The general strategy when proving something about finite-volume hyperbolic manifolds is the following: we first suppose that the manifold is compact for simplicity, and then we adapt the proof to the non-compact case by looking at what happens to the cusps.

4.3.1. Geodesic spectrum. We start by proving the following.

Proposition 4.3.1. *Let M be a finite-volume complete hyperbolic manifold. For every $L > 0$ there are finitely many closed geodesics in M shorter than L.*

Proof. Suppose that there are infinitely many closed geodesics shorter than L. We know that M decomposes into a compact part and a finite union of truncated cusps. The compact part has finite diameter D.

Every closed geodesic intersects the compact part because a cusp contains no closed geodesic. Therefore we can fix a basepoint $x_0 \in M$ and connect x_0 to these infinitely many geodesics of length $< L$ with arcs shorter than D. We use these arcs to freely homotope the geodesics into loops based at x_0 of length $< L + 2D$, and lift the loops to arcs in $\mathbb{H}^n$ starting from some basepoint $\tilde{x}_0 \in \mathbb{H}^n$.

If two such arcs end at the same point, the corresponding closed geodesics in M are freely homotopic: this is excluded by Proposition 4.1.13, hence these endpoints are all distinct. The orbit of $\tilde{x}_0$ now contains infinitely many points in the ball $B(\tilde{x}_0, L + 2D)$, a contradiction because the orbit is discrete. $\square$

The lengths of the closed geodesics in M form a discrete subset of $\mathbb{R}$ called the *geodesic spectrum* of M. Let $\ell_1(M) > 0$ be the minimum of the spectrum of M: a *shortest closed geodesic* in M is a closed geodesic of shortest length $\ell_1(M)$. It is not necessarily unique (not even if considered up to orientation reversal), but there are finitely many of them.

Proposition 4.3.2. *If M is a closed hyperbolic manifold, then*

$$\operatorname{inj} M = \frac{1}{2} \cdot \ell_1(M)$$

and every shortest closed geodesic is simple.

Proof. The length of a closed geodesic is the minimum displacement of the corresponding hyperbolic transformation, so Corollary 4.1.2 gives the equality.

Let now γ be a shortest closed geodesic. Consider Proposition 4.1.10: if γ wraps multiple times along a curve η, then η is shorter than γ, a contradiction. If γ self-intersects transversely at some point p, we can split γ naturally as $\gamma_1 * \gamma_2$ in $\pi_1(M, p)$, where both γ_1 and γ_2 are shorter than γ. Either γ_1 or γ_2 is non-trivial in $\pi_1(M, p)$ and is hence hyperbolic, but it has length smaller than $\ell_1(M)$: this contradicts Corollary 4.1.15. $\square$

Remark 4.3.3. A shortest geodesic may be non-simple when M has cusps! By doubling an ideal triangle along its boundary we construct a hyperbolic surface called the *thrice-punctured sphere* which has three cusps and contains various closed geodesics, none of which is simple. See Chapter 6.

4.3.2. Isometry group. We now study the isometry group $\operatorname{Isom}(M)$ of a hyperbolic manifold M. Recall that the *normaliser* $N(H)$ of a subgroup $H < G$ is the set of elements $g \in G$ such that $gH = Hg$. It is the biggest subgroup of G containing H as a normal subgroup. The isometry group $\operatorname{Isom}(M)$ has an algebraic representation.

Proposition 4.3.4. *Let $M = \mathbb{H}^n / _\Gamma$ be a complete hyperbolic manifold. There is a natural isomorphism*

$$\operatorname{Isom}(M) \cong N(\Gamma) / _\Gamma.$$

Proof. Every isometry $\varphi \colon M \to M$ lifts to an isometry $\tilde{\varphi}$

$$
\begin{array}{ccc}
\mathbb{H}^n & \overset{\tilde{\varphi}}{\dashrightarrow} & \mathbb{H}^n \\
\pi \downarrow & & \downarrow \pi \\
M & \overset{\varphi}{\longrightarrow} & M
\end{array}
$$

such that $\tilde{\varphi}\Gamma = \Gamma\tilde{\varphi}$: hence $\tilde{\varphi} \in N(\Gamma)$. The lift is uniquely determined up to left- or right-multiplication by elements in Γ, hence we get a homomorphism

$$\operatorname{Isom}(M) \to N(\Gamma) / _\Gamma$$

which is clearly surjective (every element in $N(\Gamma)$ determines an isometry) and injective (if $\tilde{\varphi} \in \Gamma$ then $\varphi = \mathrm{id}$). $\qquad\square$

Recall that the *centraliser* of $H < G$ is the set of elements $g \in G$ such that $gh = hg$ for all h. It is a subgroup of G.

Exercise 4.3.5. Let $M = \mathbb{H}^n/_\Gamma$ be a finite-volume hyperbolic manifold. The centralizer of Γ in $\mathrm{Isom}(\mathbb{H}^n)$ is trivial.

4.3.3. Outer automorphism group. The *automorphism group* of a group G is the group $\mathrm{Aut}(G)$ of all the isomorphisms $G \to G$. The *inner automorphisms* are those isomorphisms of type $g \mapsto hgh^{-1}$ for some $h \in G$, and they form a normal subgroup $\mathrm{Int}(G) \lhd \mathrm{Aut}(G)$. The quotient

$$\mathrm{Out}(G) = \mathrm{Aut}(G)/_{\mathrm{Int}(G)}$$

is called the *outer automorphism group* of G.

If x_0, x_1 are two points in a path-connected topological space X there is a non-canonical isomorphism $\pi_1(X, x_0) \to \pi_1(X, x_1)$, unique only up to post-composing with an inner automorphism. Therefore there is a *canonical* isomorphism $\mathrm{Out}(\pi_1(X, x_0)) \to \mathrm{Out}(\pi_1(X, x_1))$. Hence $\mathrm{Out}(\pi_1(X))$ depends very mildly on the basepoint.

The group $\mathrm{Homeo}(X)$ of all homeomorphisms of X does not act directly on $\pi_1(X)$ because of the inner-automorphism ambiguity, but we get a natural homomorphism

$$\mathrm{Homeo}(X) \longrightarrow \mathrm{Out}(\pi_1(X))$$

which is neither injective nor surjective in general. We note that homotopic self-homeomorphisms give rise to the same element in $\mathrm{Out}(\pi_1(X))$.

4.3.4. Finite isometry groups. We turn back to hyperbolic manifolds.

Proposition 4.3.6. *If M is a finite-volume complete hyperbolic manifold, the natural map*

$$\mathrm{Isom}(M) \longrightarrow \mathrm{Out}(\pi_1(M))$$

is injective.

Proof. Set $M = \mathbb{H}^n/_\Gamma$, identify Γ with $\pi_1(M)$ and $\mathrm{Isom}(M)$ with $N(\Gamma)/_\Gamma$. With these identifications the map

$$N(\Gamma)/_\Gamma \longrightarrow \mathrm{Out}(\Gamma)$$

is just the conjugacy action that sends $h \in N(\Gamma)$ to the automorphism $g \mapsto h^{-1}gh$ of Γ. This is an inner automorphism if and only if there is a $f \in \Gamma$ such that $h^{-1}gh = f^{-1}gf$ for all $g \in \Gamma$, that is if hf^{-1} commutes with g for all $g \in \Gamma$. Exercise 4.3.5 gives $h = f \in \Gamma$ and hence the map is injective. $\qquad\square$

Corollary 4.3.7. *Let M be a finite-volume complete hyperbolic manifold. Distinct isometries of M are not homotopic.*

This is a quite strong fact, from which we deduce the following.

Corollary 4.3.8. *The isometry group of every finite-volume complete hyperbolic manifold M is finite.*

Proof. We know that M is diffeomorphic to the interior of some compact N having $k \geqslant 0$ boundary components. Let $C_1, \ldots, C_k \subset M$ be disjoint truncated cusps, all of the same small volume $V > 0$. Every isometry of M permutes these truncated cusps and fixes their complement $C = M \setminus \text{int}(C_1 \cup \ldots \cup C_k)$.

The set C is compact and preserved by the Lie group $\text{Isom}(M)$, which is hence compact. To show that it is finite it suffices to prove that it is discrete. Suppose that a sequence of isometries φ_i converges to the identity: we will deduce that φ_i is homotopic to id for sufficiently large values of i, a contradiction.

For any $\varepsilon > 0$ there is a φ_i that moves all points of C at distance $< \varepsilon$. If $\varepsilon > 0$ is sufficiently small, the isometry φ_i preserves each C_i, it restricts to an isometry of the flat torus $X_i = \partial C_i$ and its extension to C_i is just a smaller rescaling of $\varphi_i|_{X_i}$ at every flat torus leaf in the cusp.

Let also $\varepsilon > 0$ be sufficiently small, so that $\varepsilon < \text{inj}_x M$ for every $x \in C$. We get $d(x, \varphi_i(x)) < \text{inj}_x M$ for all $x \in C$, and since φ_i acts on each C_i just by rescaling its action on X_i, we get the same inequality for every $x \in M$.

For every $x \in M$ there is a unique geodesic γ_x of length $d(x, \varphi_i(x))$ connecting x and $\varphi_i(x)$. The geodesics γ_x vary continuously with x and we can use them to define a homotopy between φ_i and id: a contradiction.   □

The finiteness of the isometry groups is peculiar to the hyperbolic world: we will see in Remark 7.1.3 that the isometry group of a flat torus is infinite.

4.3.5. Finite covers. Let $M = \mathbb{H}^n/_\Gamma$ be a complete finite-volume hyperbolic manifold. How many finite covers $\tilde{M} \to M$ are there above M? Quite a lot, thanks to the following.

Proposition 4.3.9. *The fundamental group $\pi_1(M)$ of a complete finite-volume hyperbolic manifold M is residually finite.*

Proof. The group $\pi_1(M)$ is finitely generated (actually, finitely presented) because M is homeomorphic to the interior of a compact manifold with boundary, so Proposition 3.1.15 applies.   □

This algebraic fact has some nice geometric consequences. Recall that $\ell_1(M)$ is the length of the shortest closed geodesic in M.

Corollary 4.3.10. *For every $L > 0$ there is a finite cover $\tilde{M} \to M$ with $\ell_1(\tilde{M}) > L$.*

Proof. There are finitely many hyperbolic elements $a_1, \ldots, a_k \in \pi_1(M)$ such that every closed geodesic in M of length $< L$ is freely homotopic

to some of these. By residually finiteness there is a finite-index normal $H \lhd \pi_1(M)$ that does not contain any of the elements $a_1, \ldots, a_k$, and hence none of their conjugates.

The finite-index H determines a finite cover $\pi \colon \tilde{M} \to M$ such that $\pi_*(\pi_1(\tilde{M})) = H$. A closed geodesic in $\tilde{M}$ cannot be shorter than L because its image in M would be freely homotopic to one of $a_1, \ldots, a_k$. $\square$

Corollary 4.3.11. *If M is a closed hyperbolic manifold, for every $R > 0$ there is a finite cover $\tilde{M} \to M$ with $\mathrm{inj}\tilde{M} > R$.*

We can summarise this by saying that every closed hyperbolic manifold has arbitrarily fat finite covers.

4.3.6. Subgroup separability. The geometric consequences of residually finiteness are quite remarkable: we now introduce a stronger algebraic notion that will lead to more geometric applications.

Let G be a group. A subgroup $H < G$ is *separable* if the intersection of all the finite-index subgroups of G containing H is H itself. In other words, for every non-trivial element $a \in G \setminus H$, there is a finite-index subgroup $G' < G$ that contains H but not a and hence "separates" H from a.

By definition, the trivial subgroup $\{e\}$ in G is separable if and only if G is residually finite. Note that, for general $H < G$, we cannot require the subgroup G' to be normal in G, because H itself may not be normal in G, as opposite to $\{e\}$. The following proposition furnishes some interesting examples.

Proposition 4.3.12. *Every maximal abelian subgroup H in a residually finite group G is separable.*

Proof. More generally, we suppose that H is maximal with respect to some word relation $f(h_1, \ldots, h_n) = e$. (In our case, $f(h_1, h_2) = [h_1, h_2]$.) Since G is residually finite, there is a sequence $N_i \lhd G$ of finite-index normal subgroups with $\cap N_i = \{e\}$.

We show that $H = \cap_i H N_i$, and we conclude since $H N_i > N_i$ has finite index in G. We have $H \subset \cap_i H N_i$. We now show that the elements of $\cap_i H N_i$ satisfy $f = e$, and by maximality of H we get $H = \cap_i H N_i$.

For every N_i and $h_1, \ldots, h_n \in H$, we have $f(h_1 N_i, \ldots, h_n N_i) \subset N_i$ since its projection in G/N_i is trivial. Therefore $\cap_i f(H N_i, \ldots, H N_i) = \{e\}$ and the elements of $\cap_i H N_i$ satisfy $f = e$. $\square$

Corollary 4.3.13. *Let M be a complete finite-volume hyperbolic n-manifold. The following subgroups are separable:*

- *the subgroup $\langle \varphi \rangle = \mathbb{Z}$ generated by a primitive hyperbolic transformation φ;*
- *the subgroup $\pi_1(T)$ generated by a $(n-1)$-torus cusp section T.*

Proof. Both subgroups are maximal abelian, see Section 4.2.1. $\square$

We deduce a couple of geometric consequences from this algebraic fact.

Corollary 4.3.14. *Let M be a complete finite-volume hyperbolic n-manifold. Every closed geodesic γ in M lifts to a closed geodesic in some finite covers of M of arbitrarily large degree.*

Proof. We have $M = \mathbb{H}^n/_\Gamma$. The closed geodesic γ is obtained by projecting the axis l of a primitive hyperbolic element $\varphi \in \Gamma$. Since $\langle \varphi \rangle$ is separable, there are subgroups $H < \pi_1(M)$ of arbitrarily large index that contain $\langle \varphi \rangle$. $\square$

We have discovered in particular that there are many closed hyperbolic manifolds of bounded injectivity radius with arbitrarily large volume. We can also use separability to promote some primitive closed geodesics to simple ones on some finite coverings.

Corollary 4.3.15. *Let M be a complete finite-volume hyperbolic n-manifold. Every primitive closed geodesic γ in M lifts to a simple closed geodesic in some finite cover of M.*

Proof. We have $M = \mathbb{H}^n/_\Gamma$. The closed geodesic γ is obtained by projecting the axis l of a primitive hyperbolic element $\varphi \in \Gamma$. If γ is not simple, it self-intersects in $k > 0$ points, which lift to k transverse intersections between l and some translates $\varphi_1(l), \ldots, \varphi_k(l)$, for some $\varphi_i \in \Gamma$. Therefore the elements $\psi \in \Gamma$ such that $\psi(l)$ intersects l transversely are those of the form $\psi = \varphi^h \varphi_i \varphi^l$ for some $i = 1, \ldots, k$ and $h, l \in \mathbb{Z}$.

Since $\langle \varphi \rangle$ is separable, there is a finite-index $H < \Gamma$ with $\varphi_1, \ldots, \varphi_k \notin H$ and $\langle \varphi \rangle \subset H$. No $\varphi^h \varphi_i \varphi^l$ lies in H, hence for every $\psi \in H$ the lines l and $\psi(l)$ coincide or are disjoint. The line l projects to a simple geodesic in $\mathbb{H}^n/_H$. $\square$

Concerning cusps, we deduce analogously the following.

Corollary 4.3.16. *Let M be a complete finite-volume hyperbolic n-manifold and $T \subset M$ be a $(n-1)$-torus cusp section. The cusp section T lifts to a cusp section in some finite covers of M with arbitrarily large degree.*

Proof. The subgroup $\pi_1(T)$ is separable, so there are subgroups $H < \pi_1(M)$ of arbitrarily large finite degree containing it. $\square$

4.4. The Bieberbach Theorem

The Margulis Lemma for Lie groups is fairly general and has important applications also in the elliptic and flat geometries. The most important one is the Bieberbach Theorem.

4.4.1. Elliptic manifolds. Recall that every complete elliptic manifold is isometric to S^n/Γ for some finite group $\Gamma < O(n) = \text{Isom}(S^{n-1})$ acting freely. We now improve the Margulis Lemma 4.2.7 for $G = O(n)$ by promoting "nilpotent" to "abelian". To this purpose we need the following.

Lemma 4.4.1. *There is a neighbourhood U of $I \in O(n)$ such that for every $A \in O(n)$ and every $B \in U$ we have*

$$[A, [A, B]] = I \implies [A, B] = I.$$

Proof. We prove the lemma for the bigger Lie group $U(n)$ consisting of all unitary matrices. An immediate computation shows that

$$[A, [A, B]] = I \iff [A, BA^{-1}B^{-1}] = I.$$

Let $\mathbb{C}^n = \oplus V_i$ be the decomposition into orthogonal eigenspaces for A and A^{-1} (with respect to the standard hermitian product of $\mathbb{C}^n$). The decomposition for the conjugate $BA^{-1}B^{-1}$ is just $\mathbb{C}^n = \oplus B(V_i)$.

Let U be the neighbourhood of I consisting of all matrices that move every vector of an angle $< \frac{\pi}{2}$. Pick $B \in U$. Since the V_i are orthogonal we get $B(V_i) \cap V_j = \{0\}$ for all $i \neq j$. On the other hand, the endomorphisms A and $BA^{-1}B^{-1}$ commute and hence have a basis of common eigenvectors: hence the only possibility is that $B(V_i) = V_i$ for all i.

The restriction $A|_{V_i}$ is just $\lambda_i I$ and hence commutes with $B|_{V_i}$, for all i. Therefore A and B commute everywhere. $\square$

Corollary 4.4.2. *There is a neighbourhood U of $I \in O(n)$ such that every finite subgroup $\Gamma < O(n)$ generated by some elements in U is abelian.*

Proof. We know from Margulis Lemma 4.2.7 that there is a U where every such Γ is nilpotent. The previous lemma promotes Γ to an abelian group, because $[A, [A, \ldots [A, B] \cdots]] = I$ implies after finitely many steps that $[A, B] = I$ for every generators $A, B \in \Gamma \cap U$. $\square$

Proposition 4.4.3. *Let G be a compact Lie group and U a neighbourhood of $e \in G$. There is a $N > 0$ such that for every group $\Gamma < G$, the subgroup $\Gamma_U < \Gamma$ generated by $\Gamma \cap U$ has index at most N in Γ.*

Proof. Let $W \subset U$ be a smaller neighbourhood such that $W^{-1} = W$ and $W^2 \subset U$ and set $N = \text{Vol}(G)/\text{Vol}(W)$ using the Haar measure for G.

We conclude by showing that if $g\Gamma_U$ and $g'\Gamma_U$ are distinct cosets of Γ_U in Γ, then $gW \cap g'W = \varnothing$. This implies that Γ_U has index at most N in Γ.

Indeed, if $gW \cap g'W \neq \varnothing$ there are $w, w' \in W$ such that $gw = g'w'$ which implies that $g^{-1}g' = w(w')^{-1} \in W^2 \subset U$ and hence $g^{-1}g' \in \Gamma_U$. $\square$

We obtain a fairly interesting corollary about finite subgroups of $O(n)$.

Corollary 4.4.4. *For every n there is a $N > 0$ such that every finite subgroup of $O(n)$ contains an abelian subgroup of index at most N.*

4.4.2. Isometries of Euclidean space. We now turn to flat manifolds. Recall that a complete flat manifold is isometric to $\mathbb{R}^n/_\Gamma$ for some discrete group $\Gamma < \mathrm{Isom}(\mathbb{R}^n)$ acting freely. Every isometry g of $\mathbb{R}^n$ can be written uniquely as

$$g \colon x \longmapsto Ax + b$$

for some $A \in \mathrm{O}(n)$ and $b \in \mathbb{R}^n$. The *rotational* and *translational part* of g are A and b respectively. Let $\mathrm{Fix}(A)$ be the fixed points of $x \mapsto Ax$.

Proposition 4.4.5. *Let $Ax + b$ be an isometry of $\mathbb{R}^n$. Then:*
- *its inverse is $A^{-1}x - A^{-1}b$,*
- *if it acts freely then $\mathrm{Fix}(A) \neq \{0\}$,*
- *$\exists$ a translation of $\mathbb{R}^n$ that conjugates it to $Ax + b'$ with $b' \in \mathrm{Fix}(A)$.*

Proof. If it acts freely then $Ax + b = x$ has no solution $x \in \mathbb{R}^n$. By rewriting the equation as $(A - I)x = -b$ we see that $A - I$ is not surjective, hence not injective, hence $\mathrm{Fix}(A)$ is non-trivial.

A translation $x + d$ conjugates $Ax + b$ into

$$\bigl(A(x + d) + b\bigr) - d = Ax + (A - I)d + b.$$

Since A is orthogonal, we get $\mathrm{Im}\,(A - I) = \ker(A - I)^\perp = \mathrm{Fix}(A)^\perp$ and hence there is a d such that $b' = (A - I)d + b \in \mathrm{Fix}(A)$. $\qquad\square$

Exercise 4.4.6. The commutator of two isometries is

$$[Ax + b, Cx + d] = [A, C]x + A(I - C)A^{-1}b + AC(I - A^{-1})C^{-1}d.$$

In particular we get

$$[Ax + b, x + d] = x + (A - I)d.$$

4.4.3. Discrete groups. The homomorphism $r \colon \mathrm{Isom}(\mathbb{R}^n) \to \mathrm{O}(n)$ that sends every isometry to its rotational part induces an exact sequence

$$0 \longrightarrow \mathbb{R}^n \longrightarrow \mathrm{Isom}(\mathbb{R}^n) \overset{r}{\longrightarrow} \mathrm{O}(n) \longrightarrow 0$$

where we indicate by $\mathbb{R}^n$ the group of translations of $\mathbb{R}^n$.

Let $\Gamma < \mathrm{Isom}(\mathbb{R}^n)$ be a discrete group. We get an exact sequence

$$0 \longrightarrow H \longrightarrow \Gamma \longrightarrow r(\Gamma) \longrightarrow 0$$

where $H \lhd \Gamma$ is the *translation subgroup* of Γ. The subgroup $r(\Gamma) < \mathrm{O}(n)$ is not necessarily discrete, as the following example shows.

Example 4.4.7. A *rototranslation* in $\mathbb{R}^3$ is a rotation of some angle θ along an axis r composed with a translation of some distance $t > 0$ in the direction of r. A rototranslation generates a discrete group $\Gamma = \mathbb{Z}$ acting freely on $\mathbb{R}^3$, whose quotient $\mathbb{R}^3/_\Gamma$ is diffeomorphic to $\mathbb{R}^2 \times S^1$. If θ is not commensurable with π the image $r(\Gamma)$ is not discrete and forms a dense subset of the circle in $\mathrm{O}(3)$ of all rotations along r.

We now extend Corollary 4.4.4 to this context.

Theorem 4.4.8. *For every n there is a N > 0 such that every discrete subgroup of* $\mathrm{Isom}(\mathbb{R}^n)$ *contains an abelian subgroup of index at most N.*

Proof. We pick a small neighbourhood $U \subset O(n)$ of I, such that the commutator is contracting in U (see the proof of the Margulis Lemma 4.2.7), it is symmetric ($U = U^{-1}$) and $A - I$ contracts vectors uniformly for all $A \in U$, more precisely $\|(A - I)v\| < \frac{1}{4}\|v\|$ for all $v \neq 0$.

Let $\Gamma < \mathrm{Isom}(\mathbb{R}^n)$ be discrete and $r(\Gamma)_U$ be the group generated by $r(\Gamma) \cap U$. By Proposition 4.4.3 the index of $r(\Gamma)_U$ in $r(\Gamma)$ is bounded by some N depending only on U. Therefore its counterimage in Γ

$$\Gamma^* = r^{-1}(r(\Gamma)_U) \cap \Gamma$$

has index in Γ bounded by N. The group Γ^* is the subgroup of Γ consisting of all $Ax + b$ with $A \in r(\Gamma)_U$. It remains to prove that Γ^* is abelian.

Let $A_1 x + b_1$ and $A_2 x + b_2$ be two elements in Γ^*. Define

$$A_{i+1}x + b_{i+1} = [A_1 x + b_1, A_i x + b_i]$$

for all $i \geqslant 2$. Exercise 4.4.6 gives

(1) $A_{i+1}x + b_{i+1} = [A_1, A_i]x + A_1(I - A_i)A_1^{-1}b_1 + A_1 A_i(I - A_1^{-1})A_i^{-1}b_i$

and hence

$$A_{i+1} = [A_1, A_i],$$
$$b_{i+1} = A_1(I - A_i)A_1^{-1}b_1 + A_1 A_i(I - A_1^{-1})A_i^{-1}b_i.$$

Suppose $A_1, A_2 \in U$. Since the commutator is contracting in U and $A_i - I$ contract vectors uniformly, we get $A_i \to I$ and $b_i \to 0$ as $i \to \infty$. Therefore $A_i x + b_i$ tends to the identity, and since Γ^* is discrete it *is* the identity for all i bigger or equal than some i_0. In particular $A_{i_0} = I$ and Lemma 4.4.1 used backwards gives $A_3 = [A_1, A_2] = I$: hence A_1 and A_2 commute.

Since $r(\Gamma^*)$ is generated by $r(\Gamma^*) \cap U$, we deduce that $r(\Gamma^*)$ is abelian. Therefore (1) may be restated as

(2) $A_{i+1}x + b_{i+1} = x + (I - A_i)b_1 + (A_1 - I)b_i.$

We now consider the case $A_1 \in U$ and $A_2 = I$. We get

$$A_{i+1}x + b_{i+1} = x + (A_1 - I)^{i-1}b_2.$$

Since $A_1 \in U$ we get $(A_1 - I)^i b_2 \to 0$ and hence $(A_1 - I)^i b_2 = 0$ for some i, which gives $(A_1 - I)b_2 = 0$ since A_1 is diagonalisable. Thus $b_2 \in \mathrm{Fix}(A_1)$.

We have proved that if Γ^* contains a translation $x + b$, then $b \in \mathrm{Fix}(A)$ for all $A \in r(\Gamma^*) \cap U$. Since these A generate $r(\Gamma^*)$, the vector b belongs to

$$W = \mathrm{Fix}(r(\Gamma^*)) = \{x \mid Ax = x \ \forall A \in r(\Gamma^*)\}.$$

Pick now two arbitrary elements $Ax + b$ and $Cx + d$ in Γ^*. By (2) we get

$$[Ax + b, Cx + d] = x + (I - C)b + (A - I)d.$$

By what just said $(I - C)b + (A - I)d \in W$. On the other hand

$$\text{Im}\,(I - C) = \ker(I - C)^{\perp} = \text{Fix}(C)^{\perp} \subset W^{\perp}$$

and hence $(I - C)b \in W^{\perp}$, and analogously $(A - I)d \in W^{\perp}$. We deduce that $(I - C)b + (A - I)d \in W \cap W^{\perp}$ is trivial and all elements in Γ^* commute. $\qquad\square$

4.4.4. Crystallographic groups. A *crystallographic group* is a discrete subgroup $\Gamma < \text{Isom}(\mathbb{R}^n)$ with compact quotient $\mathbb{R}^n/_{\Gamma}$.

Proposition 4.4.9. *The image $r(\Gamma)$ of a crystallographic group $\Gamma < \text{Isom}(\mathbb{R}^n)$ is finite.*

Proof. By Theorem 4.4.8 we may suppose that Γ is abelian. We now prove that Γ abelian implies that $r(\Gamma)$ is trivial, *i.e.* all elements in Γ are translations.

Suppose that Γ contains a non-translation $Ax + b$. We conjugate Γ by a translation as in Proposition 4.4.5 to get $b \in \text{Fix}(A)$. Pick another isometry $Cx + d$ in Γ. The commutator

$$[Ax + b, Cx + d] = x + (A - I)d - (C - I)b$$

is trivial. Since A and C commute, we get $(C - I)b \in \text{Fix}(A)$, hence $(A - I)d \in \text{Fix}(A)$ and finally $d \in \text{Fix}(A)$.

We have proved that $d \in \text{Fix}(A) \subsetneq \mathbb{R}^n$ for all elements $Cx + d$ in Γ. Therefore the Γ-orbit of $0 \in \mathbb{R}^n$ is contained in $\text{Fix}(A)$. But the compactness of $\mathbb{R}^n/_{\Gamma}$ implies that there is a compact fundamental domain, and hence a $R > 0$ such that every point in $\mathbb{R}^n$ is R-close to any fixed orbit: a contradiction since $\text{Fix}(A)$ is a proper vector subspace of $\mathbb{R}^n$. $\qquad\square$

Corollary 4.4.10. *Every crystallographic group has a finite-index translation subgroup isomorphic to $\mathbb{Z}^n$.*

Recall from Section 3.4.8 that a *flat torus* is a n-torus $\mathbb{R}^n/_{\Gamma}$ where Γ is a lattice, *i.e.* a discrete group isomorphic to $\mathbb{Z}^n$ that spans $\mathbb{R}^n$ as a vector space.

Corollary 4.4.11 (Bieberbach's Theorem). *Every closed flat n-manifold is finitely covered by a flat torus.*

We conclude the discussion by noting that there are no "cusps" in Euclidean geometry.

Proposition 4.4.12. *Every finite-volume complete flat manifold is closed.*

Proof. Let $M = \mathbb{R}^n/_{\Gamma}$ be a finite-volume flat manifold. Up to taking finite indexes, the group Γ is abelian by Theorem 4.4.8. If all the elements are translations, we are done: finite-volume easily implies that $\Gamma = \mathbb{Z}^n$, thus M is closed.

Suppose some element $Ax + b$ is not a translation: every element of Γ commutes with it and hence preserves the space $\text{Fix}(A)$ that has

dimension $k < n$; hence M is isometric to $\mathrm{Fix}(A)/_\Gamma \times \mathbb{R}^{n-k}$ and has infinite volume. $\qquad\square$

4.4.5. References. Most of the material introduced in this chapter is standard and can be found in Benedetti – Petronio [**4**], Ratcliffe [**49**], and of course in Thurston's notes [**56**]. We have also consulted Thurston's book [**57**] for the part on chrystallographic groups. Proposition 4.3.12 was taken from a paper of Long [**36**].

The sphere at infinity

We have discovered that every complete hyperbolic manifold is a quotient $M = \mathbb{H}^n/_\Gamma$ for some discrete group $\Gamma < \mathsf{Isom}(\mathbb{H}^n)$ acting freely, and we now raise the following question: how does Γ act on the boundary $\partial\mathbb{H}^n$ at infinity? Does the action of Γ on $\partial\mathbb{H}^n$ furnish some information on the geometry of M?

We show in this chapter that $\partial\mathbb{H}^n$ subdivides naturally into two Γ-invariant subsets: an open set $\Omega(\Gamma)$ called the *domain of discontinuity* where the action of Γ is properly discontinuous (like in $\mathbb{H}^n$) and a complementary closed set $\Lambda(\Gamma)$ called the *limit set* where the action of Γ is more chaotic.

We then use the limit set to define a canonical decomposition of cusped hyperbolic manifolds into ideal polyhedra, called the *Epstein–Penner decomposition*.

We also devote some time to prove Theorem 5.2.1, which states that every smooth homotopy equivalence $M \to N$ between two closed hyperbolic manifolds lifts to a map $\mathbb{H}^n \to \mathbb{H}^n$ that extends nicely to the compactifications $\overline{\mathbb{H}^n} \to \overline{\mathbb{H}^n}$. This fact will have important applications in the study of hyperbolic manifolds in the subsequent chapters.

5.1. Limit set

How does a discrete group Γ of isometries of $\mathbb{H}^n$ act on the boundary at infinity $\partial\mathbb{H}^n$? We now prove that $\partial\mathbb{H}^n$ divides canonically into two Γ-invariant subsets: an open zone $\Omega(\Gamma)$ where Γ acts properly discontinuously, and a closed one $\Lambda(\Gamma)$ where it does not. These regions are called respectively the *domain of discontinuity* and the *limit set* of Γ.

5.1.1. The limit set. Throughout this section, we let Γ be a nontrivial discrete group of isometries of $\mathbb{H}^n$. Fix a point $x \in \mathbb{H}^n$. We know that the orbit $\Gamma(x)$ is a discrete subset of $\mathbb{H}^n$. The *limit set* $\Lambda(\Gamma) \subset \partial\mathbb{H}^n$ of Γ is the set of all the accumulation points of the orbit $\Gamma(x)$ in $\overline{\mathbb{H}^n}$.

Exercise 5.1.1. The limit set does not depend on x.

The limit set is clearly a closed Γ-invariant subset of $\partial\mathbb{H}^n$. If $\Gamma' < \Gamma$ we obviously get $\Lambda(\Gamma') \subset \Lambda(\Gamma)$.

Exercise 5.1.2. If Γ' has finite index in Γ then $\Lambda(\Gamma') = \Lambda(\Gamma)$.

Recall that Γ is elementary if it fixes a finite set of points in $\overline{\mathbb{H}^n}$.

Example 5.1.3. If Γ is elementary and acts freely, that is $M = \mathbb{H}^n/_\Gamma$ is wither a cusp or a tube (see Proposition 4.2.9), then $\Lambda(\Gamma) = \mathrm{Fix}(\Gamma)$ consists of one or two points.

We now characterise the elementary subgroups of $\mathrm{Isom}(\mathbb{H}^n)$. Recall that Γ is *virtually P* if it has a finite-index subgroup that is P (where P is some property).

Proposition 5.1.4. *The following are equivalent:*
   (1) Γ *is elementary,*
   (2) Γ *fixes either a point $x \in \mathbb{H}^n$, or a line, or a point $x \in \partial\mathbb{H}^n$ and all the horospheres centred at x,*
   (3) Γ *is virtually abelian,*
   (4) $\Lambda(\Gamma)$ *consists of 0, 1, or 2 points.*

Proof. (1) $\Rightarrow$ (2). We expand on the proof of Proposition 4.2.9 by taking elliptic elements into account.

By hypothesis Γ fixes a finite set of points in $\overline{\mathbb{H}^n}$. If some of them lie in $\mathbb{H}^n$, then Γ fixes their barycenter and we are done. If Γ fixes more than two points in $\partial\mathbb{H}^n$, a barycenter is also defined: their convex hull is an ideal polyhedron of dimension at least 2, we can truncate (via horospheres) all the vertices by the same small volume and take the barycenter of the vertices of the resulting compact combinatorial polyhedron. So we are done also in this case. If Γ fixes two points, it fixes a line.

We are left with the case where Γ fixes a point $x \in \partial\mathbb{H}^n$, and no other finite set of points. The proof of Lemma 4.2.2 extends as is when φ_2 is elliptic, and shows that if Γ contains a hyperbolic element φ_1 then every other non-trivial element of Γ is either hyperbolic or elliptic and fixes the same axis of φ_1, but this is excluded: so Γ contains only parabolics and elliptics, and these must fix all the horospheres centred at x, as required.

(2) $\Rightarrow$ (3). The group Γ is virtually isomorphic to a discrete subgroup of $\mathrm{Isom}(\mathbb{R}^m)$ for some m and we apply Theorem 4.4.8.

(3) $\Rightarrow$ (4). By Exercise 5.1.2 we may suppose that Γ is abelian. By adapting the proof of Lemma 4.2.1 to elliptics we see (exercise) that Γ satisfies (2) and hence we get (4).

(4) $\Rightarrow$ (1). If $\Lambda(\Gamma)$ is empty then Γ is finite and fixes a point by Proposition 2.4.7. If it consists of one or two points, these are preserved by Γ. $\qquad\square$

5.1.2. Minimality. Let Γ be a non elementary non-trivial discrete group of isometries of $\mathbb{H}^n$. We are interested in the action of Γ on $\Lambda(\Gamma)$. We want to prove that the action is *minimal*, that is $\Lambda(\Gamma)$ has no invariant non-empty proper closed subset.

Recall from Section 3.2.1 that every closed subset $S \subset \overline{\mathbb{H}^n}$ has a well-defined convex hull $C(S) \subset \overline{\mathbb{H}^n}$. The convex hull $C(S)$ is closed and

is the intersection of all the closed half-spaces containing S. We have $C(S) \cap \partial \mathbb{H}^n = S \cap \partial \mathbb{H}^n$.

Proposition 5.1.5. *If Γ is not elementary, its action on $\Lambda(\Gamma)$ is minimal.*

Proof. Let $S \subset \Lambda(\Gamma)$ be a non-empty closed Γ-invariant subset. Since S is Γ-invariant, its convex hull $C(S)$ also is. Since Γ is not elementary, S contains at least two points and hence the intersection $C(S) \cap \mathbb{H}^n$ is not empty. If $x \in C(S) \cap \mathbb{H}^n$, its Γ-orbit is confined in $C(S)$ and hence also its accumulation points are: therefore $\Lambda(\Gamma) \subset C(S) \cap \partial \mathbb{H}^n = S$ and hence $\Lambda(\Gamma) = S$. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Corollary 5.1.6. *If Γ is not elementary and $\Gamma' \triangleleft \Gamma$ is an infinite normal subgroup, then $\Lambda(\Gamma') = \Lambda(\Gamma)$.*

Proof. For every $\gamma \in \Gamma$ we have $\gamma^{-1} \Gamma' \gamma = \Gamma'$, so γ sends the orbit $\Gamma'(x)$ to the orbit $\Gamma'(\gamma(x))$ and hence preserves $\Lambda(\Gamma')$. Therefore $\Lambda(\Gamma')$ is Γ-invariant.

The limit set $\Lambda(\Gamma')$ is not empty because Γ' is infinite: the minimality of $\Lambda(\Gamma)$ implies that $\Lambda(\Gamma') = \Lambda(\Gamma)$. $\qquad\qquad\qquad\qquad\qquad\square$

5.1.3. The convex core. If $C \subset \overline{\mathbb{H}^n}$ is a closed convex set, we define the *nearest point retraction*

$$r \colon \overline{\mathbb{H}^n} \to C$$

as the map that sends x to the point $r(x) \in C$ that is closer to x. If $x \in \partial \mathbb{H}^n$, we interpret $r(x)$ as the first point of C that is contained in some horosphere centred at x. We have $r(x) = x$ if and only if $x \in C$. The map r is continuous. Using the geodesic from x to $r(x)$ we can construct a natural deformation retraction of $\overline{\mathbb{H}^n}$ onto the closed convex set C.

Let $\Gamma < \mathrm{Isom}(\mathbb{H}^n)$ be a non elementary non-trivial discrete group. The convex hull of $\Lambda(\Gamma)$ is Γ-invariant, so it makes sense to define the following.

Definition 5.1.7. The *convex core* of the orbifold $O = \mathbb{H}^n/_\Gamma$ is the quotient $C(\Lambda(\Gamma))/_\Gamma \subset O$ of the convex hull $C(\Lambda(\Gamma))$ of the limit set $\Lambda(\Gamma)$.

We are of course mostly interested in the case where Γ acts freely and hence $M = \mathbb{H}^n/_\Gamma$ is a manifold. The deformation retraction defined above is Γ-invariant, therefore every complete hyperbolic manifold M deformation retracts onto its convex core. In particular, M is homotopically equivalent to its convex core.

5.1.4. The domain of discontinuity. Let $\Gamma < \mathrm{Isom}(\mathbb{H}^n)$ be a non elementary non-trivial discrete group. The *domain of discontinuity* of Γ is the open set

$$\Omega(\Gamma) = \partial \mathbb{H}^n \setminus \Lambda(\Gamma).$$

The following proposition explains the terminology.

Proposition 5.1.8. *The action of Γ on $\mathbb{H}^n \cup \Omega(\Gamma)$ is properly discontinuous.*

Proof. The nearest point retraction r sends $\mathbb{H}^n \cup \Omega(\Gamma)$ to $C(\Lambda(\Gamma)) \setminus \Lambda(\Gamma)$ and commutes with Γ. The action of Γ on the latter set is properly discontinuous (since it is contained in $\mathbb{H}^n$), so the action on the former also is. $\qquad\square$

Proposition 5.1.9. *If $\mathrm{Vol}(\mathbb{H}^n/_\Gamma) < +\infty$ then $\Lambda(\Gamma) = \partial\mathbb{H}^n$.*

Proof. If $\Omega(\Gamma) \neq \varnothing$, pick $x \in \Omega(\Gamma)$. The point x has a neighbourhood system consisting of half-spaces. Since Γ acts properly discontinuously on $\mathbb{H}^n \cup \Omega(\Gamma)$, there is some half-space H which intersects only finitely many Γ-translates, contradicting the finite volume hypothesis. $\qquad\square$

In this book we are mostly interested in finite-volume complete hyperbolic manifolds or orbifolds $\mathbb{H}^n/_\Gamma$, and for these the limit set is just the whole boundary, regardless of Γ. This apparently disappointing piece of information has some interesting algebraic consequences.

Corollary 5.1.10. *If $\mathrm{Vol}(\mathbb{H}^n/_\Gamma) < +\infty$ then Γ does not contain any non-trivial virtually abelian normal subgroup.*

Proof. Let $H \triangleleft \Gamma$ be virtually abelian. If H is infinite, then $\Lambda(H) = \Lambda(\Gamma) = \partial\mathbb{H}^n$ by Corollary 5.1.6, contradicting Proposition 5.1.4. If H is finite, then $\mathrm{Fix}(H)$ is a non-empty proper subspace of $\mathbb{H}^n$ and Γ acts on $\mathrm{Fix}(H)$ because H is normal. If $\mathrm{Fix}(H)$ is a point then Γ is elementary, otherwise we get $\Lambda(\Gamma) \subset \partial\mathrm{Fix}(H)$, a contradiction in both cases. $\qquad\square$

Corollary 5.1.11. *The fundamental group of a finite-volume complete hyperbolic manifold is never solvable.*

Proof. Solvable groups have non-trivial normal abelian subgroups, as proved in Proposition 1.4.10. $\qquad\square$

5.1.5. Schottky groups. In all the examples encountered up to now the limit set $\Lambda(\Gamma)$ consists of either few points or the whole of $\partial\mathbb{H}^n$. There are many cases where $\Lambda(\Gamma)$ is a more interesting (and often beautiful) set. The simplest examples are probably the following.

Choose a number $k \geqslant 2$ and $2k$ half-spaces $H_1, \ldots, H_{2k} \subset \mathbb{H}^n$ with disjoint closures in $\overline{\mathbb{H}}^n$. The closed complement $C = \mathbb{H}^n \setminus \mathrm{int}(H_1 \cup \ldots \cup H_{2k})$ is a convex subset bounded by $2k$ disjoint hyperplanes. If we pair isometrically these hyperplanes we get a complete (exercise) hyperbolic manifold M, which is hence $M = \mathbb{H}^n/_\Gamma$ for some Γ. The hyperbolic manifold M has infinite volume and is the interior of a compact manifold that decomposes into one 0-handle and k 1-handles. The group Γ is free with k generators.

Exercise 5.1.12. If $n = 2$, then $\Lambda(\Gamma) \subset \partial\mathbb{H}^2 = S^1$ is a Cantor set.

5.1.6. Parabolic and hyperbolic points. Let $\Gamma < \mathrm{Isom}(\mathbb{H}^n)$ be a non elementary non-trivial discrete group. A *parabolic point* $x \in \partial\mathbb{H}^n$ for Γ is a point that is fixed by some parabolic element $\gamma \in \Gamma$. Similarly, a *hyperbolic point* is a point that is fixed by some hyperbolic element.

Proposition 5.1.13. *The set of all parabolic (hyperbolic) points is either empty or a dense subset of $\Lambda(\Gamma)$.*

Proof. They clearly form a Γ-invariant subset, so its closure also is. We conclude because $\Lambda(\Gamma)$ is minimal. $\square$

Corollary 5.1.14. *If $\mathrm{Vol}(\mathbb{H}^n/_\Gamma) < +\infty$, the set of all parabolic (hyperbolic) points is either empty or dense in $\partial\mathbb{H}^n$.*

5.1.7. Horoballs in the hyperboloid model. In the conformal models, horoballs are just Euclidean balls tangent to the boundary at infinity. In the hyperboloid model I^n there is a more algebraic description. Let

$$L = \{x \in \mathbb{R}^{n+1} \mid \langle x, x \rangle = 0, \ x_{n+1} > 0\}$$

be the positive light cone and recall from Section 2.2.3 that the boundary at infinity ∂I^n may be interpreted as the set of rays in L. More than that, every vector $x \in L$ determines a horoball O centred at $[x] \in \partial I^n$, via the equation

$$O = \big\{y \in I^n \mid -1 \leqslant \langle x, y \rangle < 0\big\}.$$

The horoball gets smaller as x goes to infinity.

Exercise 5.1.15. This is really a horoball centred at $[x]$.

Hint. Prove that the boundary horosphere $\langle x, y \rangle = -1$ is orthogonal to all the lines pointing to x. $\square$

There is also a simple (and maybe surprising) geometric relation between the vector x and the corresponding horoball in the Poincaré disc model D^n. Represent D^n inside the hyperplane $x_{n+1} = 0$, and let $\pi \colon I^n \to D^n$ the isometry obtained by projecting towards $P = (0, \ldots, 0, -1)$, considered in Section 2.1.6. By projecting towards P we also get a homeomorphism $\pi \colon L \to D^n$.

Exercise 5.1.16. The Euclidean centre of the horoball $\pi(O)$ is $\pi(x)$.

5.1.8. The Epstein–Penner decomposition. We now show that every cusped finite-volume hyperbolic manifold decomposes canonically into some ideal polyhedra. This decomposition is known as the *canonical* or *Epstein–Penner decomposition*.

Let $M = \mathbb{H}^n/_\Gamma$ be a non-compact finite-volume complete hyperbolic manifold. The manifold M is diffeomorphic to the interior of a compact N with $c \geqslant 1$ boundary components. We fix c disjoint truncated cusps $H_1, \ldots, H_c \subset M$. Their lifts in $\mathbb{H}^n$ are c disjoint Γ-orbits $B_1^i, \ldots, B_c^i$ of disjoint horoballs.
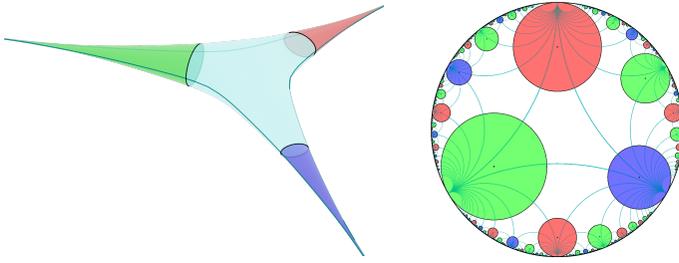
Figure 5.1. By glueing two identical ideal triangles along
their sides we get a thrice-punctured sphere S, and we choose
three disjoint truncated cusps in it having the same area
(left). In the universal cover, the two ideal triangles of S
lift to the Farey tessellation already shown in Figure 3.8, and
each of the three truncated cusps lifts to infinitely many dis-
joint horoballs (right). The Euclidean centre of the horoballs
is shown here, in relation to Exercise 5.1.16.

Example 5.1.17. Consider the thrice-punctured sphere S obtained by
gluing two identical copies of an ideal triangle along their sides. This is
a complete finite-volume hyperbolic surface with three cusps. We may
fix three disjoint truncated cusps of the same area, as sketched in Figure
5.1-(left). These lift to three families of disjoint horoballs as in Figure
5.1-(right).

We use the hyperboloid model I^n and interpret a horoball as a point in
the positive light cone L as explained in the previous section. The Γ-orbits
$B_1^i, \ldots, B_c^i$ of horoballs form a discrete subset in L. If we modify the initial
truncated cusp H_j, all the points $B_j^i \in L$ are rescaled by the same constant.

We make a crucial observation: the points B_j^i are discrete in L, but
the rays that contain them form a countable dense subset of L, because
they correspond to the parabolic points of Γ, that are dense by Corollary
5.1.14.

We now define $C \subset \mathbb{R}^{n+1}$ as the convex hull of the points B_j^i in $\mathbb{R}^{n+1}$.
We note that C is Γ-invariant. An example is shown in Figure 5.2. The
convex hull will define a decomposition of M into ideal polyhedra.

The convex hull C is contained in the convex hull of L, that consists
of all the positive timelike and lightlike vectors. The following proposition
says that C intersects L into countably many half-lines based at the B_j^i.

Proposition 5.1.18. *The set $C \cap L$ is the set of αB_j^i for some $\alpha \geqslant 1$.*

Proof. Pick a point $x \in L$ not of this type: the segment $0x$ does not
intersect $\{B_j^i\}$. Since $\{B_j^i\} \subset L$ is discrete, we can perturb the hyperplane
H tangent in x to L so that it intersects L into a small ellipsoid around $0x$,
and so that $\{B_j^i\}$ lies on the opposite side of H of x. Therefore $x \notin C$.
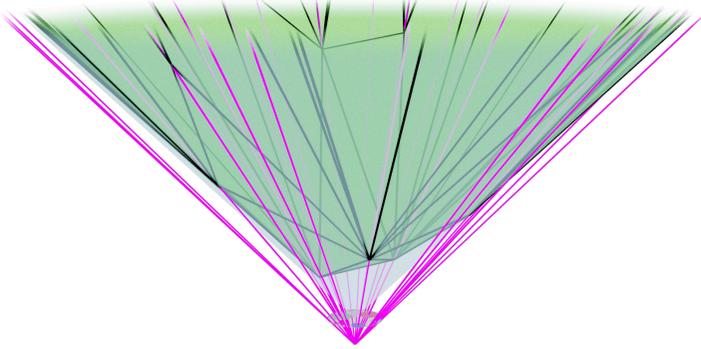
Figure 5.2. A portion of the convex hull C for the thrice-punctured sphere of Figure 5.1. As suggested by Exercise 5.1.16 we determine the horoballs in L by projecting the Euclidean centers of their representations in the disc model. The black half-line above each horoball in L is contained in C, see Proposition 5.1.18.

Conversely, pick a point $x = \alpha B_{j_0}^{i_0}$. The rays containing the points B_j^i form a dense set in L, so we can find a sequence of $B_j^i \neq B_{j_0}^{i_0}$ such that $[B_j^i]$ converges to $[B_{j_0}^{i_0}]$. Since $\{B_j^i\}$ is discrete, the points B_j^i go to infinity in the sequence and hence the segment with endpoints B_j^i and $B_{j_0}^{i_0}$ approaches x in the limit. Therefore $x \in C$. $\square$

Proposition 5.1.19. *Every timelike ray intersects ∂C exactly once.*

Proof. Every timelike ray r enters in the interior of a polyhedron spanned by some B_j^i, so it intersects C and hence ∂C. Consider a point $x \in \partial C \cap r$. Every supporting hyperplane for x must be spacelike because the rays containing the B_j^i are dense. Therefore $C \cap r$ is a half-line and x is its endpoint. $\square$

In particular $\partial C = (\partial_l C) \sqcup (\partial_t C)$ consists of the lightlike vectors $\partial_l C = C \cap L$ and some timelike vectors $\partial_t C$. The previous proposition furnishes a natural 1-1 correspondence $\partial_t C \longleftrightarrow I^n$ with the hyperboloid model by projecting along timelike lines.

Proposition 5.1.20. *The timelike boundary $\partial_t C$ is tessellated by countably many Euclidean n-dimensional polyhedra with vertices in $\{B_j^i\}$.*

Proof. Consider a point $x \in \partial_t C$ and a supporting hyperplane H. As already mentioned, the hyperplane H is spacelike and the intersection $H \cap \partial_t C$ is some k-dimensional polyhedron, convex hull of finitely many
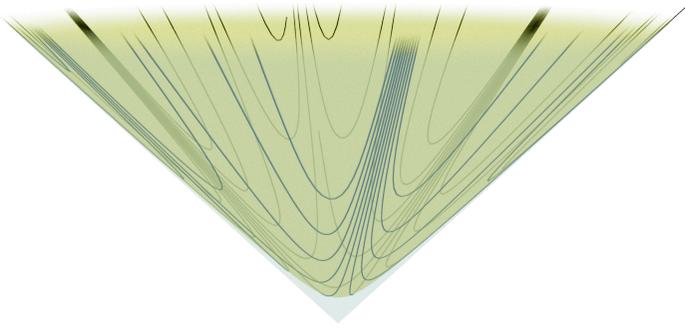
Figure 5.3. The Γ-invariant decomposition of the hyperbolic plane into ideal polygons obtained by projecting the Euclidean polygons of Figure 5.2 from ∂C to I^2. This is in fact the Farey tessellation of Figure 5.1.

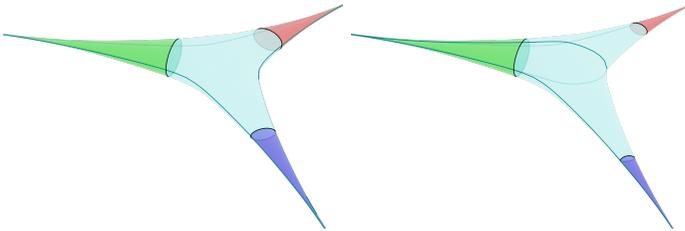Figure 5.4. Different choices of truncated cusps may lead to distinct decompositions into ideal polyhedra.

points in $\{B_j^i\}$. If $k < n$ we can rotate H until it meets one more point B_j^i and after finitely many rotations we get a n-dimensional polyhedron.

We have proved that $\partial_t C$ is paved by n-dimensional polyhedra with vertices in $\{B_j^i\}$ and intersecting in common faces. The polyhedra are locally finite because $\{B_j^i\}$ is discrete, hence they form a tessellation. $\quad\square$

The tessellation of $\partial_t C$ is Γ-invariant and projects to a Γ-invariant tessellation of the hyperbolic space I^n into ideal polyhedra, which projects in turn to a tessellation of M into finitely many ideal polyhedra, called the *canonical* or *Epstein-Penner decomposition*. An example is shown in Figure 5.3. We have discovered the following.

Theorem 5.1.21. *Every finite-volume non-compact complete hyperbolic manifold can be tessellated into finitely many ideal polyhedra.*

The canonical tessellation depends only on the chosen initial truncated cusps $H_1, \ldots, H_c$. Different choices may give different canonical tessellations, an example is sketched in Figure 5.4. We can parametrize these choices by recording the volumes $V_1, \ldots, V_c > 0$ of the truncated cusps (or the areas of ∂H_i, that are $V_i/(n-1)$ by Proposition 4.1.7).

Proposition 5.1.22. *If we multiply all volumes $V_1, \ldots, V_c$ by the same constant $\lambda > 0$ we get the same canonical tessellation.*

Proof. The set B_j^i changes by a global rescaling, and hence also the tessellation does. Its projection on I^n is unaffected. $\square$

A way to get a truly canonical decomposition of M, which depends on no choice, consists of taking $V_1 = \ldots = V_c > 0$. Note that when $c = 1$ any choice leads to the same canonical decomposition.

5.2. Extensions of homotopies

This section is entirely devoted to the proof of the following theorem. Recall that every map $f\colon M \to N$ between connected manifolds lifts to a (non-unique) map $\tilde{f}\colon \tilde{M} \to \tilde{N}$ between their universal covers.

Theorem 5.2.1. *Let $f\colon M \to N$ be a smooth homotopy equivalence between closed hyperbolic n-manifolds. Every lift $\tilde{f}\colon \mathbb{H}^n \to \mathbb{H}^n$ extends to a continuous map $\tilde{f}\colon \overline{\mathbb{H}^n} \to \overline{\mathbb{H}^n}$ whose restriction $\tilde{f}|_{\partial\mathbb{H}^n}\colon \partial\mathbb{H}^n \to \partial\mathbb{H}^n$ is a homeomorphism.*

Every smooth homotopy equivalence extends to a homeomorphism of the boundaries of the universal coverings. This theorem will have important consequences in the Chapters 8 and 13. To prove it, we need to define a weaker notion of isometry.

5.2.1. Quasi-isometries. We introduce the following.

Definition 5.2.2. A map $F\colon X \to Y$ between metric spaces is a *quasi-isometry* if there are two constants $C_1 > 0$, $C_2 \geqslant 0$ such that

$$\frac{1}{C_1}d(x_1, x_2) - C_2 \leqslant d(F(x_1), F(x_2)) \leqslant C_1 d(x_1, x_2) + C_2$$

for all $x_1, x_2 \in X$ and $d(F(X), y) \leqslant C_2$ for all $y \in Y$.

A quasi-isometry is an isometry up to some error: note that F might neither be continuous nor injective. Two metric spaces are *quasi-isometric* if there is a quasi-isometry $F\colon X \to Y$ (which implies the existence of a quasi-isometry $G\colon Y \to X$, exercise) and quasi-isometry is an equivalence relation between metric spaces. Intuitively, looking at a space up to quasi-isometries is like watching it from some distance. Compact metric spaces are obviously quasi-isometric to a point.

This notion is an important ingredient in *geometric group theory*: one may for instance give every finitely-presented group G a canonical metric (through its *Cayley graph*), uniquely determined up to quasi-isometries.

5.2.2. Pseudo-isometries. Let $f\colon M \to N$ be a homotopic equivalence between closed hyperbolic n-manifolds. Every continuous function is homotopic to a smooth one, hence we suppose that f is smooth. The map lifts to a map $\tilde{f}\colon \mathbb{H}^n \to \mathbb{H}^n$. We will prove that $\tilde{f}$ is a quasi-isometry. Actually, the map $\tilde{f}$ is also continuous and (as we soon see) Lipschitz: it will be useful for us to retain this information on $\tilde{f}$ to simplify some arguments, so we introduce a different (more restrictive but less natural) version of a quasi-isometry.

Definition 5.2.3. A map $F\colon X \to Y$ between metric spaces is a *pseudo-isometry* if there are two positive constants $C_1, C_2 > 0$ such that

$$\frac{1}{C_1} d(x_1, x_2) - C_2 \leqslant d(F(x_1), F(x_2)) \leqslant C_1 d(x_1, x_2)$$

for all $x_1, x_2 \in X$.

In particular a pseudo-isometry is C_1-Lipschitz and hence continuous. Let $f\colon M \to N$ be a smooth map between Riemannian n-manifolds; the *maximum dilatation* of f at a point $x \in M$ is the maximum ratio $\frac{\|df_x(v)\|}{\|v\|}$ where v varies among all the unitary vectors in T_x. The maximum dilatation of f is the supremum of all maximum dilatations as $x \in M$ varies.

Exercise 5.2.4. If $f\colon M \to N$ has maximum dilatation C the map f is C-Lipschitz.

Proposition 5.2.5. *Let $f\colon M \to N$ be a smooth homotopy equivalence of closed hyperbolic n-manifolds. Every lift $\tilde{f}\colon \mathbb{H}^n \to \mathbb{H}^n$ is a pseudo-isometry.*

Proof. Let $g\colon N \to M$ be a homotopic inverse. The homotopy between $g \circ f$ and id_M lifts to a homotopy between a lift $\widetilde{g \circ f}$ and $\mathrm{id}_{\mathbb{H}^n}$. Let $\tilde{g}$ be a lift of g such that $\tilde{g}(\tilde{f}(p)) = (\widetilde{g \circ f})(p)$ for some p. It follows that $\widetilde{g \circ f} = \tilde{g} \circ \tilde{f}$.

Since M is compact, the map f has some finite maximum dilatation C. Since $\tilde{f}$ is locally like f, it also has maximum dilatation C and is hence C-Lipschitz. The same holds for g. Therefore there is a $C_1 > 0$ such that

$$d\big(\tilde{f}(x_1), \tilde{f}(x_2)\big) \leqslant C_1 \cdot d(x_1, x_2) \quad \forall x_1, x_2 \in \mathbb{H}^n,$$
$$d\big(\tilde{g}(y_1), \tilde{g}(y_2)\big) \leqslant C_1 \cdot d(y_1, y_2) \quad \forall y_1, y_2 \in \mathbb{H}^n.$$

Since M is compact, the homotopy from id_M to $g \circ f$ moves every point of M along some arc with uniformly bounded length $< K$. The lifted homotopy $\tilde{g} \circ \tilde{f}$ inherits the same property, hence

$$d(x_1, x_2) - 2K \leqslant d\big(\tilde{g}(\tilde{f}(x_1)), \tilde{g}(\tilde{f}(x_2))\big) \leqslant C_1 \cdot d\big(\tilde{f}(x_1), \tilde{f}(x_2)\big)$$

for all $x_1, x_2 \in \mathbb{H}^n$. Therefore $\tilde{f}$ is a pseudo-isometry with $C_2 = 2K/C_1$. $\square$
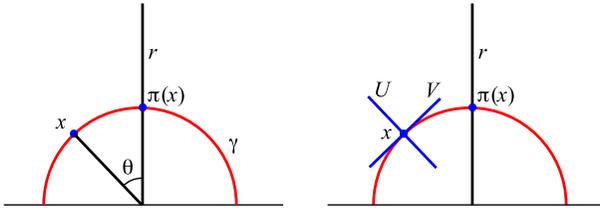
Figure 5.5. We use the half-space model. The hyperbolic
cosine of the distance between x and $\pi(x)$ is the inverse of
the cosine of θ (left). To determine the maximum dilatation
we decompose the tangent space T_x orthogonally as $U \oplus V$
(right).

5.2.3. Boundary extension of a pseudo-isometry. We now prove
that not only pseudo-isometries $\mathbb{H}^n \to \mathbb{H}^n$ extend at infinity, but their
extensions at $\partial\mathbb{H}^n$ eliminate the "errors" and behave nicely.

Theorem 5.2.6. *Every pseudo-isometry $F\colon \mathbb{H}^n \to \mathbb{H}^n$ extends to a
continuous map $F\colon \overline{\mathbb{H}^n} \to \overline{\mathbb{H}^n}$ that injects $\partial\mathbb{H}^n$ into itself.*

We separate the proof in some lemmas.

Lemma 5.2.7. *Consider the picture in Figure 5.5. We have*

$$\cosh d(x, \pi(x)) = \frac{1}{\cos \theta}.$$

Proof. We can work with the half-plane model $H^2 \subset \mathbb{C}$ and up to
translations and dilations suppose that $\pi(x) = i$. The geodesic r is
parametrized as ie^t. The Möbius transformation $z \mapsto \frac{z+1}{-z+1}$ sends r to γ
and fixes i, hence $\gamma(t) = \frac{ie^t+1}{-ie^t+1}$. Set $s = d(x, \pi(x))$. We get $x = \frac{ie^s+1}{-ie^s+1}$
and

$$\cos \theta = \Im x = \Im \frac{(ie^s + 1)^2}{e^{2s} + 1} = \frac{2e^s}{e^{2s} + 1} = \frac{2}{e^s + e^{-s}} = \frac{1}{\cosh s}.$$

The proof is complete.   □

Lemma 5.2.8. *Let $r \subset \mathbb{H}^n$ be a line and $\pi\colon \mathbb{H}^n \to r$ be the orthogonal
projection to r. The maximum dilatation of π at $x \in \mathbb{H}^n$ is*

$$d = \frac{1}{\cosh s}$$

where $s = d(x, r)$.

Proof. We use the half-space model with r and x as in Figure 5.5-
(left): we know that $\cosh s = \frac{1}{\cos \theta}$. We have $T_x = U \oplus V$ as in Figure
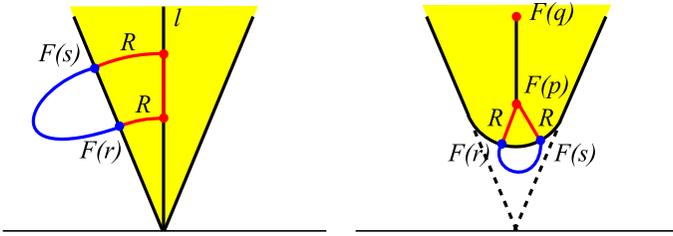5.5-(right) with $V = \ker d\pi_x$. A generator u of U is just rotated by $d\pi_x$

Figure 5.6. We use the red paths to estimate the distance between $F(r)$ and $F(s)$. On the left: since F is C_1-Lipschitz, the blue path has length at most $C_1 d(r, s)$. Its projection onto l has dilatation at most $1/\cosh R$ by Lemma 5.2.8, hence the red path in l has length at most $C_1 d(r, s)/\cosh R$. Therefore $d(F(r), F(s)) \leqslant C_1 \frac{d(r,s)}{\cosh R} + 2R$. On the right we get $d(F(r), F(s)) \leqslant 2R$.

with respect to the Euclidean metric; with respect to the hyperbolic metric we have

$$\frac{\|d\pi_x(u)\|}{\|u\|} = \frac{x_n}{\pi(x)_n} = \cos\theta = \frac{1}{\cosh s}.$$

The proof is complete. □

We write $\overline{pq}$ for the segment with endpoints p and q and we denote by $N_r(A)$ the r-neighbourhood of A.

Lemma 5.2.9. *Let* $F\colon \mathbb{H}^n \to \mathbb{H}^n$ *be a pseudo-isometry. There is a* $R > 0$ *such that*

$$F(\overline{pq}) \subset N_R\big(\overline{F(p)F(q)}\big)$$

for all distinct points $p, q \in \mathbb{H}^n$.

Proof. Let C_1, C_2 be the pseudo-isometry constants of F. Fix a sufficiently big R so that $\cosh R > 2C_1^2$. Let l be the line containing $F(p)$ and $F(q)$. We show that $F(\overline{pq})$ can exit from $N_R(l)$ only for a limited amount of time. Let $\overline{rs} \subset \overline{pq}$ be a maximal segment where $F(\overline{rs})$ is disjoint from the interior of $N_R(l)$, as the blue arc in Figure 5.6-(left). We have

$$\frac{1}{C_1} d(r, s) - C_2 \leqslant d(F(r), F(s)) \leqslant C_1 d(r, s).$$

We can improve the right inequality as shown in Figure 5.6-(left) and write

$$\frac{1}{C_1} d(r, s) - C_2 \leqslant d(F(r), F(s)) \leqslant C_1 \frac{d(r, s)}{\cosh R} + 2R.$$

Therefore

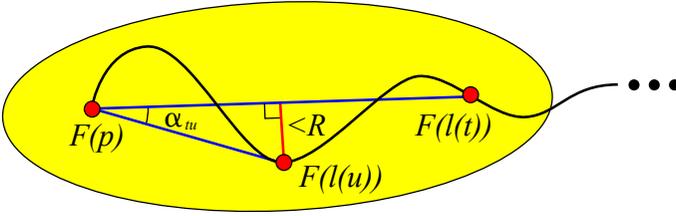$$\left(\frac{1}{C_1} - \frac{C_1}{\cosh R}\right) d(r, s) \leqslant 2R + C_2.$$

Figure 5.7. For every $0 < u < t$, the point $F(l(u))$ is con-
tained in the (yellow) R-neighbourhood of $\overline{F(p)F(l(t))}$. If u
is big, the blue segment $\overline{F(p)F(l(u))}$ is long, while the red
one is bounded by R: hence the angle α_{tu} between v_t and v_u
is small. Therefore v_t is a Cauchy sequence.

Since $\cosh R > 2C_1^2$ we deduce that $d(r, s) < M$ for some constant M
that depends only on C_1 and C_2.

We have proved that $F(\overline{pq})$ may exit from $N_R(l)$ only on subsegments
of $\overline{pq}$ with length $< M$. Since F is C_1-Lipchitz the curve $F(\overline{pq})$ lies entirely
in $N_{R+C_1 M}(l)$, and we replace R with $R + C_1 M$.

It remains to prove that $F(\overline{pq})$ lies entirely (up to taking a bigger R)
in the bounded set $N_R(\overline{F(p)F(q)})$: the proof is analogous and easier, since
Figure 5.6-(right) shows that $d(F(r), F(s)) \leqslant 2R$. $\qquad\square$

In the previous and following lemmas, the constant R depends only
on the pseudo-isometry constants C_1 and C_2.

Lemma 5.2.10. *Let $F\colon \mathbb{H}^n \to \mathbb{H}^n$ be a pseudo-isometry. There is a
$R > 0$ such that for all $p \in \mathbb{H}^n$ and every half-line l starting from p there
is a unique half-line l' starting from $F(p)$ such that*

$$F(l) \subset N_R(l').$$

Proof. We parametrize l as a geodesic $l\colon [0, +\infty) \to \mathbb{H}^n$ with unit
speed. We have $l(0) = p$. Since F is a pseudo-isometry we get

$$\lim_{t \to \infty} d\big(F(p), F(l(t))\big) = \infty.$$

Let $v_t \in T_{F(p)}$ be the unitary tangent vector pointing towards $F(l(t))$:
Figure 5.7 shows that $\{v_t\}_{t \in \mathbb{N}}$ is a Cauchy sequence, hence it converges
to a unitary vector $v \in T_{F(p)}$. Let l' be the half-line starting from $F(p)$
with direction v. It is easy to check that $F(l) \subset N_R(l')$ and l' is the unique
half-line from p with this property. $\qquad\square$

The previous lemma gives a recipe to transform every half-line l into
a half-line l' that approximates $F(l)$. Since $\partial \mathbb{H}^n$ is an equivalence relation
of half-lines, we define the extension $F\colon \partial \mathbb{H}^n \to \partial \mathbb{H}^n$ by sending l to l'.

Lemma 5.2.11. *The boundary extension $F\colon \partial \mathbb{H}^n \to \partial \mathbb{H}^n$ is well-
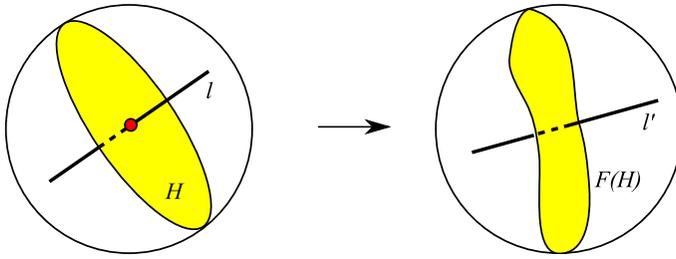defined and injective.*

Figure 5.8. Let l and H be a line and an orthogonal hyper-
plane. The orthogonal projection of H onto l is obviously a
point $l \cap H$; the pseudo-isometry F mildly distorts this picture:
the image $F(H)$ projects to a bounded segment in l'.

Proof. Let l_1, l_2 be two half-lines at bounded distance $d(l_1(t), l_2(t)) <$
M for all t. If $d(l_1'(t), l_2'(t)) \to \infty$ we get $d\big(F(l_1(t)), F(l_2(t))\big) \to +\infty$, a
contradiction since F is Lipschitz. Therefore l_1', l_2' are at bounded distance
and F is well-defined.

Injectivity is proved analogously: if l_1 and l_2 are divergent then l_1' and
l_2' also are because F is a pseudo-isometry. $\qquad\square$

It remains to prove that the extension $F \colon \overline{\mathbb{H}^n} \to \overline{\mathbb{H}^n}$ is continuous. We
start by extending Lemma 5.2.10 from half-lines to lines.

Lemma 5.2.12. *Let $F \colon \mathbb{H}^n \to \mathbb{H}^n$ be a pseudo-isometry. There is a
$R > 0$ such that for every line l there is a unique line l' with $F(l) \subset N_R(l')$.*

Proof. Parametrize l as $l \colon (-\infty, +\infty) \to \mathbb{H}^n$ with unit speed. By
cutting l into two half-lines we know that $F(l(t))$ is a curve that tends to
two distinct points $x_\pm \in \partial \mathbb{H}^n$ as $t \to \pm\infty$. Let l' be the line with endpoints
$x_\pm$. For any $t > 0$ we have

$$F\big(l([-t, t])\big) \subset N_R\left(\overline{F(l(-t))F(l(t))}\right)$$

and by sending $t \to +\infty$ we deduce that $F(l) \subset N_R(l')$. $\qquad\square$

The next lemma says that a pseudo-isometry does not distort much
lines and orthogonal hyperplanes. We will need it to prove continuity.

Lemma 5.2.13. *Let $F \colon \mathbb{H}^n \to \mathbb{H}^n$ be a pseudo-isometry. There is a
$R > 0$ such that for any line l and hyperplane H orthogonal to l, the image
$F(H)$ projects orthogonally to l' onto a bounded segment length smaller
than R.*

Proof. See Figure 5.8. Consider a generic line $s \subset H$ passing through
$p = l \cap H$. By the previous lemmas $F(s) \subset N_R(s')$ with $s' \neq l'$, and
the orthogonal projection on l' sends any other line s' onto a segment,
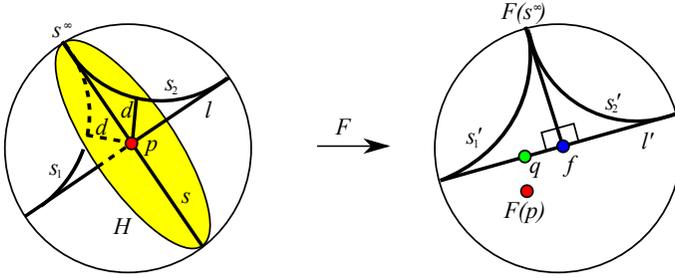bounded by the images of the endpoints of s'.

Figure 5.9. The lines s_1 and s_2 have a distance d from p which depends on nothing (in fact, $\cosh d = \sqrt{2}$). The lines l', s_1', and s_2' approximate up to an error R the images of l, s_1, and s_2 along F. The projection q of $F(p)$ on l' is hence R-close to $F(p)$, which is in turn $(C_1 d)$-close to the lines s_i'. Therefore q is $(C_1 d + 2R)$-close to both s_1' and s_2'. This easily implies that f is $(C_1 d + 2R)$-close to q.

Consider as in Figure 5.9 the line s, with one endpoint s^∞ and the corresponding endpoint $F(s^\infty)$ of s'. The figure shows that the projection f of $F(s^\infty)$ to l' is at bounded distance from a point q which does not depend on s. □

Finally, we prove that the extended F is continuous.

Lemma 5.2.14. *The extension* $F \colon \overline{\mathbb{H}^n} \to \overline{\mathbb{H}^n}$ *is continuous.*

Proof. Consider $x \in \partial \mathbb{H}^n$ and its image $F(x) \in \partial \mathbb{H}^n$. Let l be a half-line pointing to x: hence l' points to $F(x)$. The half-spaces orthogonal to l' determine a neighbourhood system for $F(x)$: consider one such half-space S.

Let $R > 0$ be as in the previous lemmas. The image $F(l)$ is R-close to l', hence for sufficiently big t the point $F(l(t))$ and its projection into l' lie in S at distance $> R$ from ∂S. By the previous lemma the image $F(H(t))$ of the hyperplane $H(t)$ orthogonal to l in $l(t)$ is also contained in S. Hence the entire half-space bounded by one such $H(t)$ goes inside S through F. This shows that F is continuous at every point $x \in \partial \mathbb{H}^n$. □

With some effort, we have proved that every pseudo-isometry of $\mathbb{H}^n$ extends continuously to the boundary. Theorem 5.2.1 now follows easily.

Proof of Theorem 5.2.1. We know that $\tilde{f}$ is a pseudo-isometry and hence extends to a map $\tilde{f} \colon \overline{\mathbb{H}^n} \to \overline{\mathbb{H}^n}$ that sends injectively $\partial \mathbb{H}^n$ to itself. It remains to prove that $\tilde{f}|_{\partial \mathbb{H}^n}$ is a homeomorphism.

Pick a smooth homotopic inverse g for f. The homotopy $\mathrm{id}_M \sim g \circ f$ lifts to a homotopy $\mathrm{id}_{\mathbb{H}^n} \sim \tilde{g} \circ \tilde{f}$ for some lift $\tilde{g}$. Since M is compact,

the latter homotopy moves every point at uniformly bounded distance and hence $\tilde{g} \circ \tilde{f}$ extends continuously to the identity on $\partial \mathbb{H}^n$, and the same holds for $\tilde{f} \circ \tilde{g}$. Therefore $\tilde{g}|_{\partial \mathbb{H}^n}$ is the inverse of $\tilde{f}|_{\partial \mathbb{H}^n}$ and they are both homeomorphisms.   □

5.2.4. References. As in the previous chapters, we have mostly consulted Benedetti – Petronio [**4**], Ratcliffe [**49**], and Thurston's notes [**56**]. The Epstein – Penner decomposition is taken from their 1988 paper [**17**].

Part 2

Surfaces

Surfaces

A surface is a differentiable manifold of dimension $n = 2$, possibly with boundary. The closed orientable surfaces are classified topologically by their Euler characteristic, a complete invariant that also determines their possible geometries: a closed surface has a hyperbolic, flat, or spherical structure if and only if its Euler characteristic is negative, null, or positive.

We devote some time here to expose the topological classification and the geometrisation of closed surfaces. Then we describe some of the beautiful features of geometrisation: the geometry of a surface (typically, hyperbolic geometry) can be used to prove in an elegant way various non-trivial topological facts.

We end this chapter by defining and studying the *mapping class group*, a group that encodes the topological symmetries of a surface.

6.1. Topological classification

A surface can be topologically quite complicated: think for instance of $\mathbb{R}^2$ with a Cantor set removed. We decide to restrict our investigation to the surfaces *of finite type*, i.e. obtained from a closed one by removing points and/or open discs: these include all the compact surfaces with or without boundary. We introduce here these surfaces and classify them up to diffeomorphism.

We manipulate surfaces using various cut-and-paste tools: boundary gluings, removal of discs or points, handle decompositions, connected sums. See Section 1.1 to refresh these notions.

6.1.1. Gluing surfaces. A simple way to construct a surface is by gluing simpler surfaces along their boundaries. We show here that the glued surface depends only on the orientation classes of the gluing maps.

Two self-diffeomorphisms of S^1 are *co-oriented* if they both preserve or both invert the orientation of S^1.

Lemma 6.1.1. *Co-oriented self-diffeomorphisms of S^1 are isotopic.*

Proof. Let $f_0, f_1 \colon S^1 \to S^1$ be two co-oriented self-diffeomorphisms. Their lifts $\tilde{f}_0, \tilde{f}_1 \colon \mathbb{R} \to \mathbb{R}$ are periodic and monotone, hence $\tilde{f}_t = (1-t)\tilde{f}_0 + t\tilde{f}_1$ also is and descends to an isotopy f_t connecting f_0 and f_1. $\qquad \square$

Figure 6.1. A surface of genus 3 may be represented in various ways.

Corollary 6.1.2. *If we glue two oriented surfaces along their boundaries via orientation-reversing diffeomorphisms, the resulting oriented surface does not depend on the diffeomorphisms chosen.*

Proof. All the orientation-reversing gluing maps are isotopic by Lemma 6.1.1, and isotopic gluing maps produce diffeomorphic manifolds, as stated in Proposition 1.1.15.   □

In a non-oriented context, we have two possible maps for every glued boundary component.

Corollary 6.1.3. *If we glue two discs we get a sphere.*

Proof. The two discs are copies of $D^2 \subset \mathbb{C}$ glued along a diffeomorphism $\varphi\colon S^1 \to S^1$. Up to mirroring one we may suppose that φ is orientation-reversing, and by Corollary 6.1.2 we may suppose that $\varphi(z) = \bar{z}$. The resulting surface is diffeomorphic to a sphere (exercise).   □

Warning 6.1.4. In dimension $n \geqslant 7$, by gluing two discs we may get a manifold that is homeomorphic but *not* diffeomorphic to a sphere! See Section 6.4.2.

6.1.2. Classification of surfaces. In Section 1.1.13 we introduced the connected sum, a two-steps operation which consists of first removing balls and then gluing the new sphere boundaries. Let S_g be the connected sum

$$S_g = \underbrace{T \# \ldots \# T}_{g}$$

of g tori $T = S^1 \times S^1$. By convention $S_0 = S^2$ is the sphere and $S_1 = T$ is the torus. The number g is the *genus* of the closed surface S_g. The surface S_g may be represented in $\mathbb{R}^3$ in various ways, see Figure 6.1.

Proposition 6.1.5. *We have $\chi(S_g) = 2 - 2g$.*

Proof. Let S, S' be surfaces and $D \subset \mathrm{int}(S), D' \subset \mathrm{int}(S')$ discs. Then

$$\chi(S \# S') = \chi(S \setminus D) + \chi(S' \setminus D') - \chi(S^1) = \chi(S) + \chi(S') - 2.$$

Therefore $\chi(S_g) = 2 - 2g$ by induction on g.   □
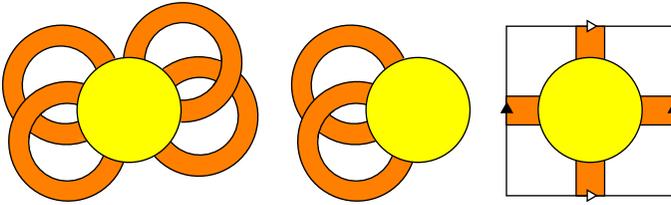
Figure 6.2. The 0-handle and k 1-handles form a subsurface $S' \subset S$ with connected boundary, to which the 2-handle is attached (left). A 0-handle and two linked 1-handles (centre). If we attach a 2-handle to the centre figure we get a handle decomposition of the torus (right).

Theorem 6.1.6 (Classification of surfaces). *Every closed, connected, orientable surface is diffeomorphic to S_g for some $g \geqslant 0$.*

Proof. Being a closed manifold, every closed orientable surface S has a handle decomposition. By Proposition 1.7.13 it has one with one 0-handle, a certain number k of 1-handles, and one 2-handle. We get $\chi(S) = 2 - k$. We prove by induction on k that $k = 2g$ is even and S is diffeomorphic to S_g.

If $k = 0$ then S is obtained by gluing two discs (the 0- and 2-handle) and is hence a sphere by Corollary 6.1.3.

Suppose $k > 0$. The 0-handle is a disc and the 1-handles are rectangles attached to its boundary as in Figure 6.2-(left). Note that since S is orientable every rectangle is attached without a twist, otherwise it would create a Möbius strip. The 0- and 1-handles altogether form a compact surface $S' \subset S$ with only one boundary component, to which the 2-handle is attached.

Since $\partial S'$ is connected, every rectangle is linked to some other rectangle as in Figure 6.2-(centre). A pair of linked rectangles form a sub-subsurface $S'' \subset S' \subset S$ as in Figure 6.2-(centre) with connected boundary. If we cut S along the curve $\partial S''$ and then cap off with two discs we perform the inverse of a connected sum.

Therefore $S = S_1 \# S_2$, where S_1 is S'' with a disc attached, *i.e.* a torus as Figure 6.2-(right) shows. The surface S_2 decomposes into a 0-handle, $k - 2$ 1-handles, and one 2-handle. We conclude by induction on k. □

6.1.3. Homology. The homology of S_g is easily calculated.

Proposition 6.1.7. *We have*

$$H_0(S_g, \mathbb{Z}) = \mathbb{Z}, \quad H_1(S_g, \mathbb{Z}) = \mathbb{Z}^{2g}, \quad H_2(S_g, \mathbb{Z}) = \mathbb{Z}.$$
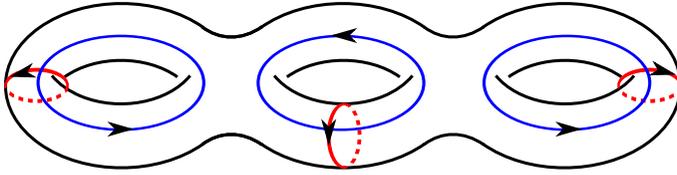
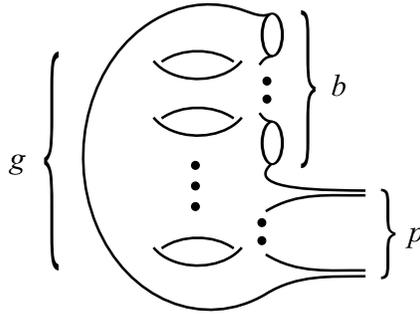Figure 6.3. These 6 oriented curves represent a symplectic basis for $H_1(S_3, \mathbb{Z}) = \mathbb{Z}^6$.

Figure 6.4. The finite type surface $S_{g,b,p}$.

Proof. Since S_g is closed, connected, and orientable, we have $H_0(S_g) = H_2(S_g) = \mathbb{Z}$. Since $\chi(S_g) = 2 - 2g$, the group $H_1(S_g)$ has rank $2g$. By Poincaré duality $H_1(S_g) = H^1(S_g) = \mathrm{Hom}(\pi_1(S_g), \mathbb{Z})$ has no torsion. □

Recall from Section 1.6.3 that by fixing an orientation for S_g we get a symplectic intersection form ω on $H_1(S_g, \mathbb{Z})$. A basis for $H_1(S_g, \mathbb{Z})$ is *symplectic* if $\omega = \begin{pmatrix} 0 & I_g \\ -I_g & 0 \end{pmatrix}$ with respect to this basis: see Figure 6.3.

The *algebraic intersection* of two closed curves is the intersection form of their classes in $H_1(S_g, \mathbb{Z})$; if the curves are transverse, this is just the algebraic sum of their intersections, where each intersection counts as ± 1 according to the local orientations.

6.1.4. Surfaces of finite type. We extend our investigation to a larger interesting class of surfaces.

Definition 6.1.8. Let $g, b, p \geqslant 0$ be three natural numbers. The *surface of finite type* $S_{g,b,p}$ is the surface obtained from S_g by removing the interior of b disjoint discs and p points.

See Figure 6.4. We say that $S_{g,b,p}$ has genus g, has b boundary components, and p punctures. Its Euler characteristic is

$$\chi(S_{g,b,p}) = 2 - 2g - b - p.$$

We also use the notation $S_{g,b}$ to indicate $S_{g,b,0}$.

Proposition 6.1.9. *Every compact connected orientable surface with boundary is diffeomorphic to $S_{g,b}$ for some $g, b \geqslant 0$.*

Proof. Let S be a compact orientable surface with some b boundary components. If we glue b discs to ∂S we get a closed orientable surface, hence diffeomorphic to S_g for some $g \geqslant 0$. The original S is obtained from S_g by removing the interiors of b disjoint open discs. □

The compact connected orientable surfaces with $\chi > 0$ are $S^2 = S_0$ and $D^2 = S_{0,1}$, while those with $\chi = 0$ are the annulus $A = S_{0,2}$ and the torus $T = S_1$.

Corollary 6.1.10 (Smooth Jordan curve Theorem). *Every smooth simple closed curve $\gamma \subset \mathbb{R}^2$ bounds a disc.*

Proof. Consider the curve γ inside $S^2 = \mathbb{R}^2 \cup \{\infty\}$. By cutting S^2 along γ we get one or two compact orientable surfaces with non-empty boundary and with $\chi = 2$ in total. The only possibility is that we get two discs. □

6.1.5. Triangulations. Like every honest smooth compact manifold, compact surfaces admit smooth triangulations, see Section 1.7.7. Conversely, we may use simplicial complexes to construct smooth surfaces combinatorially.

Let X be a two-dimensional pure simplicial complex where every edge is incident to two faces, and the link of every vertex is a circle.

Proposition 6.1.11. *The complex X is the smooth triangulation of a closed surface S, unique up to diffeomorphism.*

Proof. By dualising X we get a handle decomposition: triangles, edges, and vertices transform into 0-, 1-, and 2-handles, and we get a smooth surface S triangulated by X. The way the handles are attached is determined up to isotopy, therefore S is determined up to diffeomorphism. □

Warning 6.1.12. It is worth noting that this procedure (getting a unique smooth structure from a simplicial complex) does not work in all dimensions (here we used implicitly Lemma 6.1.1).

6.2. Geometrisation

We now prove that every surface $S_{g,b,p}$ of finite type can be *geometrised*, that is it may be equipped with a hyperbolic, flat, or elliptic metric. The metric type is prescribed by the sign of the Euler characteristic of the surface.
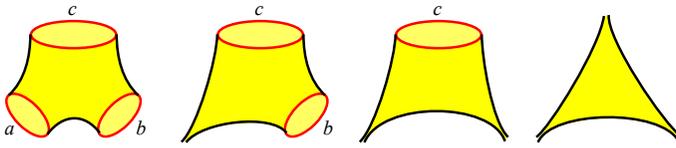
Figure 6.5. A pair-of-pants, an annulus with one puncture, a disc with two punctures, and a thrice-punctured sphere. The last three surfaces may be considered as some degenerated hyperbolic pairs-of-pants where one or more boundary lengths a, b, or c are zero, and we get cusps instead of geodesic boundary components there.

6.2.1. Hyperbolic pair-of-pants. The sphere has of course an elliptic structure, and the torus has many flat structures, see Section 3.4.8. We now construct hyperbolic structures on all the surfaces S_g of genus $g \geqslant 2$, and more generally on all the surfaces $S_{g,b,p}$ of negative Euler characteristic. We start with a simple block, the *pair-of-pants* $S_{0,3}$, with Euler characteristic -1.

Proposition 6.2.1. *Given three real numbers $a, b, c \geqslant 0$ there is (up to isometries) a unique complete finite-volume hyperbolic pair-of-pants with geodesic boundary, with boundary curves of length a, b, and c.*

When some length in a, b, c is zero, we mean that the geodesic boundary is actually a cusp (recall Section 4.2.5) and hence the surface is topologically a punctured annulus $S_{0,2,1}$, a twice punctured disc $S_{0,1,2}$, or a thrice-punctured sphere $S_{0,0,3}$: see Figure 6.5.

To prove this proposition we construct some right-angled hexagons in $\mathbb{H}^2$ as in Figure 6.6-(left). Three *alternate sides* on a hexagon are three pairwise non-incident sides, like the a, b, c shown in the figure. A *degenerate* hexagon is one where the length of some alternate sides is zero as in Figure 6.7.

Lemma 6.2.2. *Given three real numbers $a, b, c \geqslant 0$ there exists (up to isometries) a unique (possibly degenerate) hyperbolic right-angled hexagon with three alternate sides of length a, b, and c.*

Proof. We first suppose $a, b > 0$. The construction of the hexagon is depicted in Figure 6.6-(right). If $x = 0$ the blue lines coincide, hence $P = Q$ and $f(0) = 0$. The function $f : [0, +\infty) \to [0, +\infty)$ is continuous, strictly monotonic, and with $\lim_{x \to \infty} f(x) = \infty$: therefore there is precisely one x such that $f(x) = c$.

If exactly two parameters are zero, say $a = b = 0$, a simpler construction works: take a segment of length c as in Figure 6.7-(centre), draw the perpendiculars at their endpoints, and a line connecting the endpoints of these.
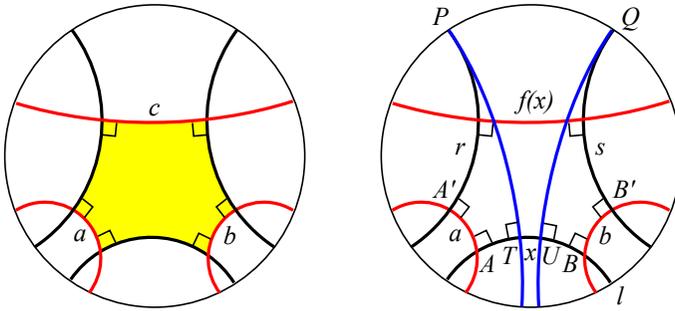
Figure 6.6. A right-angled hexagon with alternate sides of length a, b and c (left) and its construction (right), which goes as follows: take a line l with two arbitrary points A and B in it (bottom black). Draw the perpendiculars from A and B (red). At distances a and b we find two points A' and B' and we draw again two perpendiculars (black) r and s, with some points at infinity P and Q. Draw the (unique) perpendiculars to l pointing to P and Q (blue): they intersect l in two points T and U. Note that AT and UB have some fixed length depending only on a and b. We can vary the parameter $x = TU$: if $x > 0$ the lines r and s are ultra-parallel and there is a unique segment orthogonal to both of some length $f(x)$.
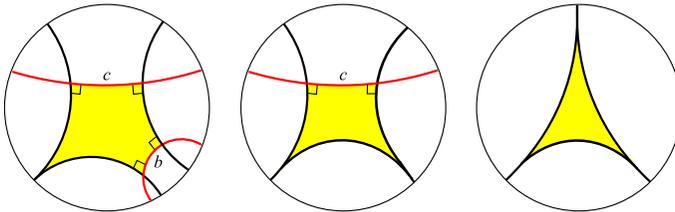
Figure 6.7. A right-angled hexagon with parameters $a, b, c \geqslant 0$ degenerates to a pentagon, quadrilateral, or tri-angle with ideal vertices if one, two, or three parameters are zero.

If $a = b = c = 0$, pick any ideal triangle. Ideal triangles are indeed unique up to isometry: use the half-space model and recall that $\mathbb{PSL}_2(\mathbb{R})$ acts transitively on the ordered triples of points in $\partial\mathbb{H}^2$.   □

The most degenerate case is so important that we single it out.

Corollary 6.2.3. *All the ideal triangles in $\mathbb{H}^2$ are isometric.*
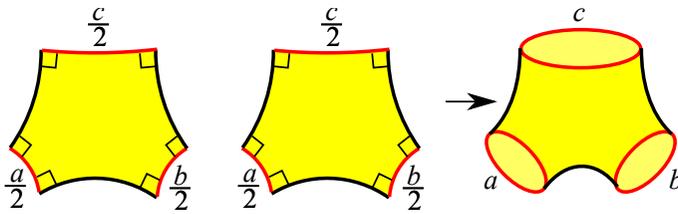
Figure 6.8. By gluing two identical right-angled hexagons along their black sides we get a hyperbolic pair-of-pants with geodesic boundary.

By gluing two identical (possibly degenerate) hexagons along alternate sides as in Figure 6.8 we construct a (possibly degenerate) hyperbolic pair-of-pants whose geodesic boundary consists of three simple closed geodesics of length $2a$, $2b$, and $2c$.

Proof of Proposition 6.2.1. We have proved the existence of a geodesic pairs-of-pants with any parameters $a, b, c \geqslant 0$, and we now turn to its uniqueness. Let P be a pair-of-pants with geodesic boundary curves C_1, C_2, C_3 of length $a, b, c > 0$.

Since P, C_1, and C_2 are compact, there are points $x_1 \in C_1$ and $x_2 \in C_2$ at minimum distance $d = d(x_1, x_2)$, connected by some curve γ_3 of length d. The curve is a simple geodesic orthogonal to both C_1 and C_2: if not, some other curve connecting x_1 and x_2 would be shorter. We construct analogously two orthogeodesics γ_1 and γ_2 connecting C_2 to C_3 and C_3 to C_1 having minimal length.

The fact that $\gamma_1, \gamma_2, \gamma_3$ have minimal length easily implies that they are disjoint (if they intersect, we find shorter curves). The three orthogeodesics subdivide P into two hexagons, with alternate sides of length $L(\gamma_1)$, $L(\gamma_2)$, and $L(\gamma_3)$: by Lemma 6.2.2 the two hexagons are isometric, and hence the three other alternating sides also have the same length $\frac{a}{2}, \frac{b}{2}$, and $\frac{c}{2}$. Hexagons are unique up to isometry and hence the original pair-of-pants also are.

We can extend the argument to the more general case $2a, 2b, 2c \geqslant 0$ as follows. If $a = 0$, a neighbourhood of the puncture is a cusp, and we truncate it at some small horocycle C_1 (we do the same if $b = 0$ or $c = 0$). After these truncations we get a compact pair-of-pants (whose boundary is *not* geodesic at the horocycles) and decompose it into two hexagons as above. The resulting curve γ_3 is orthogonal to the horocycle C_1, hence it extends to a half-line pointing towards the puncture. The curves $\gamma_1, \gamma_2, \gamma_3$ decompose the surface into degenerate hexagons. $\square$

6.2.2. Hyperbolic surfaces. The pairs-of-pants can be used as building blocks to construct topologically all finite type surfaces with $\chi < 0$.
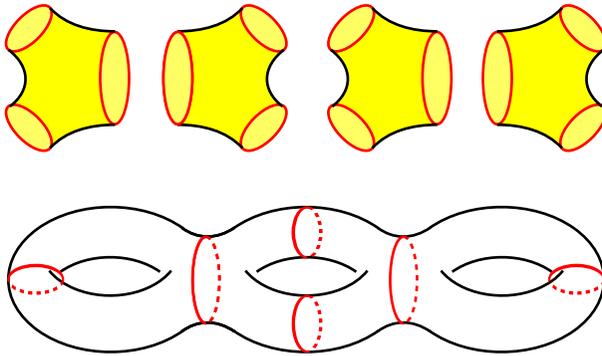
Figure 6.9. Every surface of finite type with $\chi < 0$ decomposes into pair-of-pants. We show here a decomposition of S_3.

Proposition 6.2.4. *If $\chi(S_{g,b,p}) < 0$ then $S_{g,b,p}$ decomposes topologically into $-\chi(S_{g,b,p})$ (possibly degenerate) pairs-of-pants.*

Proof. If $b + p = 0$ then $g \geqslant 2$ and the surface decomposes easily in many ways, see for instance Figure 6.9. If $b + p > 0$ and $\chi < -1$, a decomposition for $S_{g,b,p}$ may be obtained from one of $S_{g,b-1,p}$ or $S_{g,b,p-1}$ by inserting one more (possibly degenerate) pair-of-pants. If $\chi = -1$ the surface is either a pair-of-pants, or a torus with a puncture or boundary component, which is in turn obtained by glueing two boundary components of a pair-of-pants.   □

We can use this building block to construct hyperbolic structures.

Corollary 6.2.5. *If $\chi(S_{g,b,p}) < 0$ then $S_{g,b,p}$ admits a complete hyperbolic metric with b geodesic boundary components of arbitrary length.*

Proof. Decompose $S_{g,b,p}$ in pair-of-pants, assign an arbitrary length to all the closed curves of the decomposition (the 6 red curves shown in Figure 6.9) and give each pair-of-pants the hyperbolic metric determined by the three assigned boundary lengths. Everything glues to a complete hyperbolic metric for $S_{g,b,p}$ thanks to Proposition 3.5.4.   □

Can we geometrise the few orientable surfaces with $\chi \geqslant 0$? Yes, but since there are no cusps in the elliptic and flat geometries we do not consider surfaces with punctures. The compact orientable surfaces with $\chi > 0$ are the sphere and the disc, and they all have an elliptic metric with geodesic boundary (represent the disc as a hemisphere). Those with $\chi = 0$ are the torus and the annulus, and they admit flat metrics with geodesic boundary.
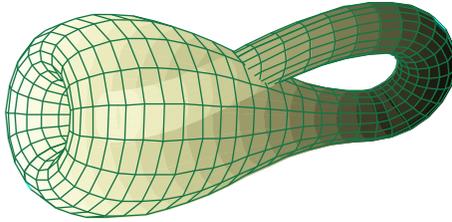
Figure 6.10. The Klein bottle immersed in $\mathbb{R}^3$.

6.2.3. Non-orientable surfaces. The classification of all the finite-type non-orientable surfaces is also simple. Let S_g^{no} be the connected sum

$$S_g^{\text{no}} = \underbrace{\mathbb{R}\mathbb{P}^2 \# \ldots \# \mathbb{R}\mathbb{P}^2}_{g}$$

of $g \geqslant 1$ copies of the projective plane $\mathbb{R}\mathbb{P}^2$.

Proposition 6.2.6. *We have $\chi(S_g^{\text{no}}) = 2 - g$. The surface S_2^{no} is diffeomorphic to the Klein bottle K shown in Figure 6.10. We have*

$$S_g \# \mathbb{R}\mathbb{P}^2 \cong S_{2g+1}^{\text{no}}.$$

Proof. The formula for χ follows from $\chi(\mathbb{R}\mathbb{P}^2) = 1$. The Klein bottle K may be cut along a closed curve into two Möbius strips, and $\mathbb{R}\mathbb{P}^2$ minus an open disc is a Möbius strip too, hence $\mathbb{R}\mathbb{P}^2 \# \mathbb{R}\mathbb{P}^2 = K$. The latter equality is a consequence of $T \# \mathbb{R}\mathbb{P}^2 \cong K \# \mathbb{R}\mathbb{P}^2$ which is left as an exercise. $\square$

Proposition 6.2.7. *Every closed, connected, non-orientable surface is diffeomorphic to S_g^{no} for some $g \geqslant 1$.*

Proof. Pick a handle decomposition of the surface S. Since it is non-orientable, at least one 1-handle is twisted and forms a Möbius strip. We have proved that S contains a Möbius strip, and we now remove it and substitute it with a disc to get a new surface S'. We have $S = S' \# \mathbb{R}\mathbb{P}^2$ and we conclude by induction on $-\chi(S')$. $\square$

We may also denote by $S_{g,b,p}^{\text{no}}$ the surface obtained from S_g^{no} by removing the interiors of b discs and p points.

Exercise 6.2.8. If $\chi(S_{g,b,p}^{\text{no}}) < 0$ then $S_{g,b,p}^{\text{no}}$ decomposes into pairs-of-pants and hence admits a complete hyperbolic metric with geodesic boundaries of arbitrary length.

As in the orientable case, compact surfaces with $\chi \geqslant 0$ can be geometrised: these are $\mathbb{R}\mathbb{P}^2$, the Möbius strip, and the Klein bottle. The first has an elliptic structure while the other two have flat structures with geodesic boundary.
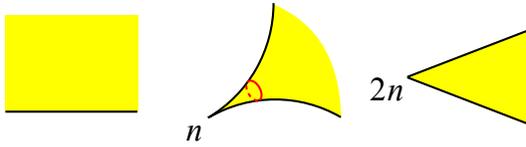
Figure 6.11. A mirror, a cone point, and a corner reflector
on a two-dimensional orbifold.

6.2.4. Orbifolds. We can push the classification and geometrisation
further to compact two-dimensional orbifolds.

The finite subgroups of $O(2)$ are: the cyclic $\mathbb{Z}_2$ generated by a re-
flection, the cyclic $\mathbb{Z}_n$ generated by a rotation, and the dihedral D_{2n} con-
taining reflections and rotations. Therefore every singular point x on a
two-dimensional orbifold O is locally of one of these types:

$$V/_{\mathbb{Z}_2}, \quad V/_{\mathbb{Z}_n}, \quad V/_{D_{2n}}$$

with V the unit ball in $\mathbb{R}^2$. The point x is called respectively a *mirror*, a
cone, and a *corner reflector* point of O, see Figure 6.11.

We consider for simplicity only *locally orientable* orbifolds, *i.e.* orbifolds
with isotropy groups in $SO(2)$. In other words, we exclude mirrors and
corner reflectors. One such orbifold O is easily encoded as

$$(S, p_1, \ldots, p_k)$$

where S is a surface, and the orbifold O is S with k cone points with
rotational isotropy groups of order $p_1, \ldots, p_k > 1$. We do not require O
to be globally orientable, so S can be a non-orientable surface like $\mathbb{RP}^2$.
We define the *Euler characteristic* of that orbifold as

$$\chi(O) = \chi(S) - \sum \left(1 - \frac{1}{p_i}\right).$$

This definition is designed to behave well under coverings. The *degree* of
an orbifold covering $O \to O'$ is the cardinality of the fibre of any non-
singular point in O'.

Proposition 6.2.9. *If $O \to O'$ is a degree-d orbifold covering then*

$$\chi(O) = d \cdot \chi(O').$$

Proof. Pick a cellularisation of O' whose vertices contain the cone
points: it lifts to a similar cellularisation of O. We consider all the vertices
of these cellularisations as cone points, possibly with order 1, and we get

$$\chi(O') = V - E + F - \sum \left(1 - \frac{1}{p'_i}\right) = -E + F + \sum \frac{1}{p'_i}$$

orbifold	Γ	orbifold	Γ
S^2	$\{e\}$	$(S^2, 2, 2, n)$	D_{2n}
(S^2, n, n)	$\mathbb{Z}_n$	$(S^2, 2, 3, 3)$	T_{12}
$\mathbb{RP}^2$	$\mathbb{Z}_2$	$(S^2, 2, 3, 4)$	O_{24}
$(\mathbb{RP}^2, n)$	D_{2n}	$(S^2, 2, 3, 5)$	I_{60}

Table 6.1. Except the two bad cases, every locally orientable closed orbifold O with $\chi(O) > 0$ can be geometrised as $O = S^2/_\Gamma$ for some discrete $\Gamma < O(3)$. In all cases $\Gamma < SO(3)$ and hence O is orientable, except $\mathbb{RP}^2$ and $(\mathbb{RP}^2, n)$ whose Γ contains the antipodal map. The groups on the right table are the spherical Von Dyck groups, see Section 3.6.4. The groups T_{12}, O_{24}, I_{60} are the orientation-preserving isometry groups of the tetrahedron, octahedron, and icosahedron, and are isomorphic to A_4, S_4, A_5 respectively.

where V, E, F is the number of vertices, edges, and faces in the cellularisation of O'. The same formula holds for O:

$$\chi(O) = -dE + dF + \sum \frac{1}{p_j}.$$

A cone point of order p_i' in O' lifts to some cone points with orders $p_{j_1}, \ldots, p_{j_l}$ of O, such that

$$d = \sum_{a=1}^{l} \frac{p_i'}{p_{j_a}} \quad \Longrightarrow \quad d \cdot \frac{1}{p_i'} = \sum_{a=1}^{l} \frac{1}{p_{j_a}}.$$

This implies that $\chi(O) = d \cdot \chi(O')$. $\square$

6.2.5. Geometrisation of orbifolds. Recall that an orbifold is good if it is covered by a manifold. With a couple of exceptions, all the closed two-orbifolds are good and geometric.

Theorem 6.2.10. *Every closed locally orientable 2-orbifold O is good except the following*

$$(S^2, p), \quad (S^2, p_1, p_2)$$

with $p_1 \neq p_2$. A good orbifold O has an elliptic, flat, hyperbolic structure $\Longleftrightarrow \chi(O)$ is positive, zero, negative. The elliptic orbifolds are listed in Table 6.1.

Proof. The orbifolds (S^2, p) and (S^2, p_1, p_2) are bad: subdivide S^2 into two discs, each containing at most one cone point; each disc has a unique surface (disc) covering, but the two coverings do not match if $p \neq 1$ or $p_1 \neq p_2$. The remaining orbifolds with $\chi(O) > 0$ are

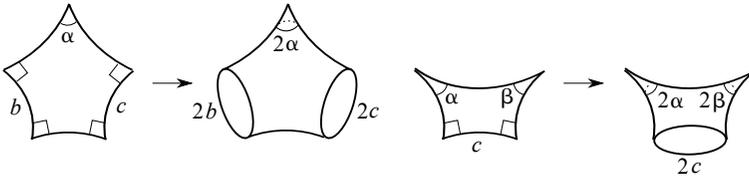$$S^2, \ (S^2, n, n), \ \mathbb{RP}^2, \ (\mathbb{RP}^2, n),$$

Figure 6.12. For every angle $\alpha \leqslant \frac{\pi}{2}$ and lengths $b, c > 0$ there is a pentagon as shown (left); by doubling it along the unmarked edges we get a hyperbolic annulus with a cone point of angle 2α. For every angles $\alpha, \beta \leqslant \frac{\pi}{2}$ and every length $c > 0$ there is a quadrilateral as shown (right). By doubling it along the unmarked edges we get a hyperbolic disc with cone points of angle 2α and 2β.

$$(S^2, 2, 2, n),\ (S^2, 2, 3, 3),\ (S^2, 2, 3, 4),\ (S^2, 2, 3, 5)$$

and they can all be realised as S^2/Γ for an appropriate finite $\Gamma < O(3)$, see Table 6.1. The orbifolds with $\chi(O) = 0$ are

$$(S^2, 2, 3, 6),\ (S^2, 2, 4, 4),\ (S^2, 3, 3, 3),\ (S^2, 2, 2, 2, 2),$$

$$(\mathbb{RP}^2, 2, 2),\ K,\ T$$

where K and T are the Klein bottle and the torus. The first four orbifolds were obtained as $\mathbb{R}^2/\Gamma$ in Section 3.6.4. The orbifold $(\mathbb{RP}^2, 2, 2)$ is $\mathbb{R}^2/\Gamma$ where Γ is generated by two glide reflections with orthogonal axis:

$$(x, y) \mapsto (x + 1, -y), \quad (x, y) \mapsto (-x, y + 1).$$

To prove this, consider the fundamental domain $[-\frac{1}{2}, \frac{1}{2}] \times [-\frac{1}{2}, \frac{1}{2}]$.

If $\chi(O) < 0$ we construct a hyperbolic metric as we did in the surface case. If $O = (S^2, p_1, p_2, p_3)$ we get $O = \mathbb{H}^2/\Gamma^{\mathrm{or}}(p_1, p_2, p_3)$ using Von Dyck groups, see Section 3.6.4. Otherwise, the orbifold O easily decomposes along disjoint simple closed curves (that avoid the cone points) into some basic pieces with $\chi < 0$, which are of the following kind:

$$P,\ (A, p),\ (D, p_1, p_2)$$

where P, A, D are the pair-of-pants, the annulus, and the disc, and $(p_1, p_2) \neq (2, 2)$. We can give a hyperbolic structure to P with any fixed length at the boundaries.

Analogously, we can give a cone manifold structure with geodesic boundary to (A, p) (respectively, (D, p_1, p_2)) with one (two) cone point of angle $\frac{2\pi}{p}$ ($\frac{2\pi}{p_1}$ and $\frac{2\pi}{p_2}$), for any fixed lengths at the boundaries. We already know this for P, and the proof for (A, p) and (D, p_1, p_2) is similar: instead of constructing right-angled hexagons, we construct pentagons with angles $\frac{\pi}{2}, \frac{\pi}{2}, \frac{\pi}{2}, \frac{\pi}{2}, \frac{\pi}{p}$ and quadrilaterals with angles $\frac{\pi}{2}, \frac{\pi}{2}, \frac{\pi}{p_1}, \frac{\pi}{p_2}$, and we double them, see Figure 6.12.

By gluing all the pieces we obtain a hyperbolic cone structure on O with cone angles $\frac{2\pi}{p_i}$ and we apply Proposition 3.6.18 to translate it to a hyperbolic orbifold structure.   □

Recall that an orbifold is very good if it is finitely covered by a manifold. In the following corollary we use the geometrisation to prove shortly a non-trivial topological fact.

Corollary 6.2.11. *Every good closed locally orientable 2-orbifold is also very good.*

Proof. It is geometric, so Corollary 3.6.19 applies.   □

The punctured case is also interesting when $\chi(O) < 0$. We define a *punctured orbifold* to be one of type

$$O = (S_{g,0,b}, p_1, \ldots, p_k).$$

Theorem 6.2.12. *Every punctured orbifold O with $\chi(O) < 0$ has a hyperbolic structure.*

Proof. Same proof as above, where we use triangular groups with possibly ideal vertices and we allow the pieces P, (A, p), (D, p_1, p_2) to have boundary components and/or cone points that degenerate to cusps. To construct them we build hexagons, pentagons, quadrilaterals with some lengths or angles that may be zero.   □

These punctured orbifolds are also very good for the same reason above. It is also possible to consider orbifolds based on surfaces $S_{g,b,p}$ with boundary: every boundary component can be considered in two natural but distinct ways, either as a mirror (hence the orbifold is locally non-orientable), or as a boundary component: to do this we need to define an appropriate notion of orbifold with boundary.

6.2.6. Gauss–Bonnet formula. How can we compute the area of a hyperbolic surface? We can easily answer this question in the closed case.

Proposition 6.2.13 (Gauss–Bonnet). *Let S be a closed hyperbolic surface. We have*

$$\text{Area}(S) = -2\pi\chi(S).$$

Proof. Pick a Dirichlet domain D for S. It is a $(2n)$-gon with inner angles $\alpha_1, \ldots, \alpha_{2n}$. By Proposition 3.2.9 we have

$$\text{Area}(S) = \text{Area}(D) = (2n - 2)\pi - \sum_{i=1}^{2n} \alpha_i.$$

The surface S is obtained from D by gluing isometrically some pairs of edges. We get a cellularisation of S with some v vertices, n edges, and one face D. Therefore

$$\chi(S) = v - n + 1.$$

Each of the v vertices is obtained by identifying some of the vertices of D, whose inner angles must sum to 2π. Therefore we get

$$\sum_{i=1}^{2n} \alpha_i = 2\pi v$$

and hence

$$\text{Area}(S) = (2n-2)\pi - 2\pi v = -2\pi(v - n + 1) = -2\pi\chi(S).$$

The proof is complete. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

We can extend the formula to orbifolds and to the other geometries.

Corollary 6.2.14. *Let O be a closed hyperbolic, flat, or elliptic locally orientable 2-orbifold. We have*

$$K \cdot \text{Area}(O) = 2\pi\chi(O)$$

where $K = -1$, 0, or $+1$ respectively.

Proof. If O is an orientable surface, then the equality holds: in the hyperbolic case this is Proposition 6.2.13, in the flat case we get $0 = 0$, and in the elliptic case $S = S^2$ and $\text{Area}(S^2) = 4\pi = 2\pi\chi(S^2)$.

In general there is a degree-d covering $S \to O$ of an orientable geometric surface S by Corollary 6.2.11. The equality holds for S and hence also for O since

$$\chi(S) = d \cdot \chi(O), \qquad \text{Area}(S) = d \cdot \text{Area}(O).$$

The proof is complete. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

6.2.7. Lie subgroups of $\text{SO}(3)$**.** The following proposition is of general interest and is used in many different contexts. It shows, among other things, that orientable elliptic orbifolds are *rigid*: their geometry is fully determined by their fundamental group. (Flat and hyperbolic surfaces are not rigid, as we will soon see.)

Proposition 6.2.15. *Every non-trivial proper Lie subgroup of $\text{SO}(3)$ is conjugate to one of the following:*

$$C_n, \ D_{2n}, \ T_{12}, \ O_{24}, \ I_{60}, \ \text{SO}(2), \ \text{SO}(2) \rtimes C_2.$$

These are the orientation-preserving isometry groups of: a regular n-pyramid, n-prism, tetrahedron, octahedron, icosahedron, cone, and cylinder.

Proof. The Lie algebra $\mathfrak{g} = \mathfrak{so}(3, \mathbb{R})$ is generated by the matrices

$$A_x = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}, \quad A_y = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}, \quad A_z = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

with the relations

$$[A_x, A_y] = A_z, \quad [A_y, A_z] = A_x, \quad [A_z, A_x] = A_y.$$

Therefore $\mathfrak{g}$ is isomorphic to $\mathbb{R}^3$ equipped with the standard vector product $\times$. We deduce that the only non-trivial proper sub-algebras of $\mathfrak{g}$ have dimension 1, and hence every non-trivial proper Lie subgroup $G < SO(3)$ has dimension 0 or 1.

If a Lie subgroup $G < SO(3)$ has dimension zero, it is finite. Every non-trivial element in G is a rotation along some axis, and hence it acts on S^2 with two (antipodal) fixed points. Let $P \subset S^2$ be the set of the fixed points of all the non-trivial elements of G.

The group G clearly acts on P, and a double-counting argument gives

$$2\big(|G| - 1\big) = \sum_{p \in P} \big(|G_p| - 1\big)$$

where $G_p < G$ is the stabiliser of p. The fundamental theorem on group actions says that

$$|G_p| = \frac{|G|}{|O(p)|}$$

where $O(p)$ is the orbit of p. Therefore

$$2\big(|G| - 1\big) = \sum_{p \in P} \left(\frac{|G|}{|O(p)|} - 1 \right) = \sum_{O} \big(|G| - |O|\big)$$

where in the latter we sum on orbits O. We divide by $|G|$ to get

$$(3) \qquad\qquad 2 - \frac{2}{|G|} = \sum_{i=1}^{r} \left(1 - \frac{1}{a_i} \right)$$

where $a_1, \ldots, a_r$ are the orders of the stabilisers of the r orbits.

If $r = 1$ we get $a_1 = |G| = 1$ and hence G is trivial, a contradiction. If $r = 2$ we get $a_1 = a_2 = |G|$ and hence every orbit is a single point, with stabiliser G. Therefore the points are antipodal and $G = C_n$.

If $r = 3$ then (a_1, a_2, a_3) is a triple with $\frac{1}{a_1} + \frac{1}{a_2} + \frac{1}{a_3} > 1$, that is one of

$$(2, 2, n), \quad (2, 3, 3), \quad (2, 3, 4), \quad (2, 3, 5).$$

In each case one calculates $|G|$ using (3) and deduces the cardinalities of the three orbits. For instance in the $(2, 2, n)$ case we have $|G| = 2n$ and the orbits have cardinality $(n, n, 2)$. One deduces that G contains an index-two cyclic group C_n and is the symmetry group D_{2n} of the prism.

In the $(2, 3, 3)$ case we get $|G| = 12$ and hence the orbits have order $(6, 4, 4)$. The stabiliser of a point in the order-4 orbit O is a $\frac{2\pi}{3}$-rotation that rotates the other three points: one deduces that the points in O are the vertices of a regular tetrahedron and G is its symmetry group. The other cases are treated analogously.

If G has dimension one, it contains the group $SO(2)$ of all rotations along some axis r. Either $G = SO(2)$, or there is a $g \in G \setminus SO(2)$. If $g(r) \neq r$, then G contains another $SO(2)$ of rotations along $g(r)$, and these two copies of $SO(2)$ generate a Lie subgroup of dimension $\geqslant 2$ in G, a contradiction. Therefore every element in $G \setminus SO(2)$ preserves

r and is hence a π-rotation along an axis perpendicular to r. Therefore $G = \text{SO}(2) \rtimes C_2$. □

6.3. Curves on surfaces

It is natural to study a manifold by examining the lower-dimensional sub-manifolds that it contains: we now look at closed curves in surfaces and unveil an unexpectedly rich world. We will discover that the study of curves in surfaces is tightly connected to hyperbolic geometry: each of the two topics seems designed to give us a better understanding of the other.

We first prove various facts on curves, and then we use this knowledge to deduce some topological non-trivial results on surfaces. We will end up by showing that two self-diffeomorphisms of a closed surface S_g are homotopic if and only if they are isotopic, a theorem that will be important also in our topological study of three-manifolds in the subsequent chapters.

We concentrate for simplicity on the genus-g closed orientable surfaces S_g, although much of the discussion could be easily extended to finite-type surfaces $S_{g,b,p}$ with the appropriate modifications.

6.3.1. Definitions. We start by recalling and fixing some definitions. A *curve* on a differentiable manifold M is a smooth map $\gamma \colon I \to M$ defined on some interval I, while a *closed curve* is a smooth map $\gamma \colon S^1 \to M$.

A (possibly closed) curve γ is *regular* if $\gamma'(t) \neq 0$ for all t. All the curves will be tacitly assumed to be regular; moreover, with a little abuse we will sometimes indicate by γ the *support* of the curve γ, that is its image.

A curve is *simple* if it is an embedding, and the support of a simple closed curve is a one-dimensional submanifold of M diffeomorphic to S^1. Recall that all isotopies are smooth by assumption, see Section 1.1.9. Two simple closed curves are isotopic if and only if they are ambiently isotopic. Two simple closed curves with the same support and orientation are isotopic by Lemma 6.1.1.

6.3.2. Simple closed curves on the sphere. We would like to classify the simple closed curves on a given closed surface up to isotopy, but we will soon realise that this task is harder than one could expect. A closed surface may contain many complicated closed curves, difficult to draw and visualise.

For the moment we content ourselves with a very simple case. A simple closed curve on a surface S is *trivial* if it bounds a disc.

Proposition 6.3.1. *All simple closed curves in S^2 are isotopic.*

Proof. Every simple closed curve in S^2 is trivial by Jordan's Theorem (see Corollary 6.1.10).

Theorem 1.1.14 and Lemma 6.1.1 together imply that there are at most two isotopy classes of trivial simple closed curves on any connected

surface S, depending on their orientations. When $S = S^2$ the situation is quite special and the two classes reduce to one: a π-rotation along a horizontal axis transforms an eastward-run equator into a westward-run one.   □

The simple closed curves in a torus are much more interesting.

6.3.3. Simple closed curves in the torus. We now classify the simple closed curves in the torus $T = S^1 \times S^1$ up to isotopy. The fundamental group $\pi_1(T) = \mathbb{Z} \times \mathbb{Z}$ is abelian, hence a closed curve γ is determined up to free homotopy by its class $(m, n) \in \mathbb{Z} \times \mathbb{Z}$.

Proposition 6.3.2. *The class $(m, n) \neq (0, 0)$ is represented by a simple closed curve if and only if m and n are coprime. In that case, the simple closed curve is unique up to isotopy.*

Proof. We visualise the torus as a quotient $T = \mathbb{R}^2/_{\mathbb{Z}^2}$. If m and n are coprime, the vector line generated by $(m, n) \in \mathbb{R}^2$ projects to a simple closed curve in T representing the class (m, n), see Figure 6.13.

Conversely, let $\gamma \subset T$ be a simple closed curve. We cut T along γ and get a surface S with $\chi(S) = 0$. If γ separates, then by the classification of surfaces the only possibility is that S consists of a one-holed torus and a disc, so γ is trivial and $(m, n) = (0, 0)$, which is excluded. If γ does not separate, there is another curve η intersecting it transversely in one point. This implies that η and γ have algebraic intersection ± 1 and hence (m, n) are coprime: if $(m, n) = k(p, q)$ with $k \geqslant 2$ the algebraic intersection would be divided by k.

Finally, we pick two simple closed curves γ, η of the same type $(m, n) \neq (0, 0)$ and prove that they are isotopic. We put γ, η in transverse position, and we cut T along γ to get an annulus A. If η does not intersect γ, by cutting A along η we get two annuli: hence η and γ cobound an annulus and we can use this annulus to build an isotopy between them. If η intersects γ, the curve η decomposes into arcs in A as in Figure 6.14-(left). Since the algebraic intersection of η and γ is zero, there is an arc with both endpoints in the same boundary component of A. This arc forms a bigon, which we can slide by isotopy as in Figure 6.14-(right) to decrease the intersection points in $\gamma \cap \eta$ and conclude by induction.   □

Simple closed curves in surfaces of higher genus are more complicated to classify, but are still very important. A couple of techniques used in the previous proof (cutting along curves and simplifying bigons) will be employed again in the higher-genus context.

6.3.4. Preliminaries on simple curves. Let S_g be a closed orientable surface of some genus $g \geqslant 1$. We now prove some topological facts on simple closed curves in S_g, sometimes employing hyperbolic geometry.
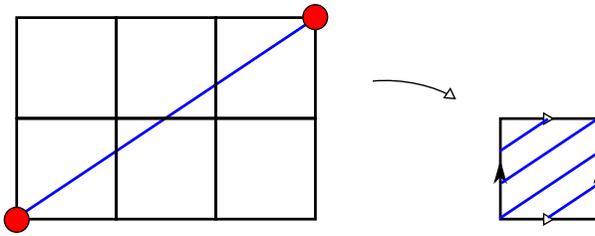
Figure 6.13. If (m, n) are coprime the line generated by $(m, n) \in \mathbb{R}^2$ projects to a simple closed curve in T. Here $(m, n) = (3, 2)$.
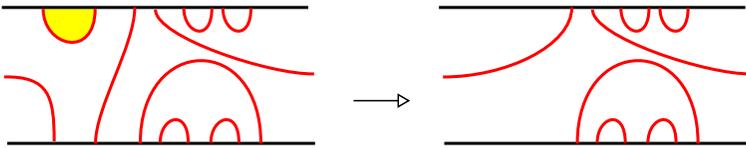

Figure 6.14. Two homotopic simple closed curves in a torus are isotopic: put them in transverse position, cut the first to get an annulus, then remove bigons to destroy intersections.

Proposition 6.3.3. *There are finitely many simple closed curves in S_g up to diffeomorphisms of S_g.*

Proof. By cutting S_g along a simple closed curve γ we get a surface S' with the same Euler characteristic as S_g, with one or two components, and with the boundary oriented as γ: there are only finitely many diffeomorphism types for S'.

Suppose that by cutting S_g along γ_1 and γ_2 we get two surfaces S_1' and S_2' of the same type. By hypothesis there is a diffeomorphism $\varphi \colon S_1' \to S_2'$ that preserves the boundary orientations. By Lemma 6.1.1 we may modify φ near the boundary so that it extends to a diffeomorphism $\varphi \colon S_g \to S_g$ sending γ_1 to γ_2. This concludes the proof. $\square$

Exercise 6.3.4. There is precisely *one* non-separating simple closed curve in S_g up to diffeomorphism!

Every non-separating simple closed curve in S_g may be transformed into your favourite one by some diffeomorphism of S_g, see Figure 6.15. This quite useful fact, analogous to the possibility of changing a basis in a vector space, was called the *change of coordinates principle* by Farb and Margalit.

Recall that a non-trivial element $g \in G$ in a group is *primitive* if it cannot be written as $g = h^n$ for some $n \geqslant 2$ and some $h \in G$. This condition is conjugacy-invariant, so the following makes sense.
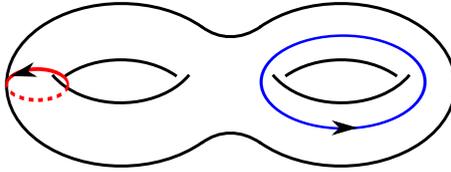
Figure 6.15. There are infinitely many non-separating simple closed curves in S_g up to isotopy when $g \geqslant 1$, but there is only one up to self-diffeomorphism of S_g. For instance, some diffeomorphism of S_2 sends the blue simple closed curve to the red one (a self-diffeomorphism may be quite hard to visualise directly...).

Proposition 6.3.5. *Let γ be a simple closed curve in S_g. It holds:*

- *if γ is homotopically trivial, it bounds a disc;*
- *if γ is not homotopically trivial, it is primitive in $\pi_1(S_g)$.*

Proof. Let S' be the surface obtained by cutting $S = S_g$ along γ. The surface S' may have one or two components and has the same Euler characteristic of S. If one component of S' is a disc, we are done. If S' is an annulus, then S is a torus and we are done by Proposition 6.3.2.

In all the other cases there is a hyperbolic metric on S where γ is a geodesic: each component of S' has negative Euler characteristic and hence can be given a hyperbolic structure with geodesic boundary curves of length 1; by gluing them we get the hyperbolic metric on S.

We now apply Proposition 4.1.13 multiple times. Since γ is a closed geodesic, it is not homotopically trivial. If $\gamma = \eta^k$ is not primitive in $\pi_1(S_g)$, then η is homotopic to a closed geodesic $\bar{\eta}$ and hence γ is homotopic to $\bar{\eta}$ run k times: a closed curve cannot be homotopic to two distinct closed geodesics, a contradiction. □

We have employed hyperbolic geometry to prove a topological fact on surfaces: this will be a refrain in this chapter. Let the *inverse* γ^* of a closed curve γ be γ run with opposite orientation.

Proposition 6.3.6. *A non-trivial simple closed curve in S_g is never freely homotopic to its inverse.*

Proof. If $g = 1$ the curves γ and γ^* represent distinct elements (and hence conjugacy classes) in $\pi_1(S_1) = \mathbb{Z} \times \mathbb{Z}$. If $g \geqslant 2$, give S_g a hyperbolic metric. The curve γ is homotopic to a closed geodesic $\bar{\gamma}$ and hence γ^* is homotopic to its inverse $\bar{\gamma}^*$, which is distinct from $\bar{\gamma}$ as a closed geodesic by definition. Distinct closed geodesics are not homotopic. □

6.3.5. Simple closed geodesics. Let now S_g have genus $g \geqslant 2$ and be equipped with a hyperbolic metric. We know from Corollary 4.1.14 that
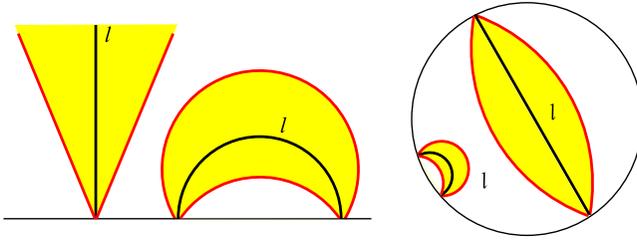
Figure 6.16. The cones and bananas R-neighbourhoods of a geodesic l in the half-plane (left) and disc (right) models.

every homotopically non-trivial closed curve in S_g has a unique geodesic representative. We now prove that, if the original curve is simple, then the geodesic representative also is.

Recall that the *R-neighbourhood* of an object in a metric space is the set of all points of distance at most R from that object. The R-neighbourhoods of lines in $\mathbb{H}^2$ are particularly simple.

Proposition 6.3.7. *The R-neighbourhood of a line $l \subset \mathbb{H}^2$ in a conformal model is bounded by two Euclidean lines or circle arcs having the same endpoints as l as in Figure 6.16.*

Proof. Put l in the half-space model with endpoints at 0 and ∞. A R-neighbourhood is invariant by the isometry $x \mapsto \lambda x$ and is hence a cone as in the figure. The other cases follow because isometries and inversions send lines and circles to lines and circles. $\square$

We use R-neighbourhoods to prove the following.

Proposition 6.3.8. *Let S_g be equipped with a hyperbolic metric. Every non-trivial simple closed curve in S_g is homotopic to a simple closed geodesic.*

Proof. Every non-trivial simple closed curve γ in $S_g = \mathbb{H}^2/_\Gamma$ is homotopic to a closed geodesic $\bar\gamma$ by Corollary 4.1.14, and we now prove that $\bar\gamma$ is simple. The counterimage of γ in $\mathbb{H}^2$ consists of disjoint simple *arcs*, while the counterimage of $\bar\gamma$ consists of straight *lines*: we prove that these lines are also disjoint.

The homotopy between γ and $\bar\gamma$ lifts to a homotopy between the arcs and the lines. The homotopy between γ and $\bar\gamma$ has compact support, hence there is a $R > 0$ such that every point in the arcs is moved in the lifted homotopy to some distance smaller than R. Therefore every arc is contained in the R-neighbourhood of a line as in Figure 6.17-(left).

This shows that lines and arcs have the same endpoints in $\partial\mathbb{H}^2$. If two lines intersects, their endpoints are linked in the circle $\partial\mathbb{H}^2$ and hence also the corresponding arcs intersect, see Figure 6.17-(right): a contradiction.
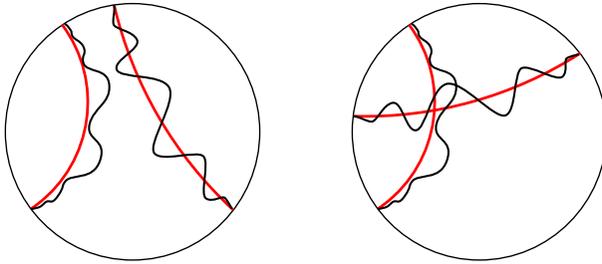
Figure 6.17. The lifts of γ (black arcs) and of its geodesic representative (red lines) have the same endpoints in $\partial \mathbb{H}^2$ (left). If two lines intersect, the arcs do (right).

Since the lifts of $\bar{\gamma}$ do not intersect, the geodesic $\bar{\gamma}$ does not self-intersect transversely. Then $\bar{\gamma}$ is either simple or wraps multiple times a simple geodesic, but the second possibility is excluded by Proposition 6.3.5. □

We will soon promote "homotopic" to "isotopic".

6.3.6. Geometric intersection. Isotopy classes of simple closed curves on S_g form a complicate and interesting set, and a way to study it consists of looking at the way these curves intersect each other. The *algebraic* intersection is too weak a tool, because it detects only their homology classes. We now introduce the much finer *geometric* intersection.

Let γ_1 and γ_2 be two simple closed curves in an orientable surface S. Let the *geometric intersection* $i(\gamma_1, \gamma_2)$ be the minimum number of intersections of two transverse simple closed curves γ_1', γ_2' homotopic to γ_1 and γ_2. The geometric intersection depends only on the homotopy classes of γ_1 and γ_2.

Proposition 6.3.9. *We have $i(\gamma, \gamma) = 0$ for every simple closed γ.*

Proof. A tubular neighbourhood of γ is diffeomorphic to $S^1 \times [-1, 1]$ because S is orientable, hence γ has two disjoint parallel representatives $S^1 \times \left\{-\frac{1}{2}\right\}$ and $S^1 \times \left\{\frac{1}{2}\right\}$ which do not intersect. □

Recall that the *algebraic* intersection of two curves counts the intersections with sign. Geometric and algebraic intersections behave much differently and are equal only modulo 2.

The geometric intersection $i(\gamma_1, \gamma_2)$ remains unaffected if we substitute γ_1 with its inverse, that is if we reverse its orientation.

6.3.7. The bigon criterion. Two simple closed curves γ_1 and γ_2 in an orientable surface S are in *minimal position* if they intersect transversely in exactly $i(\gamma_1, \gamma_2)$ points. How can we know if two transverse simple
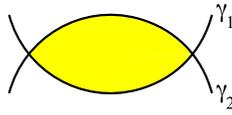
Figure 6.18. Two curves γ_1 and γ_2 are in minimal position if and only if they do not create bigons, like this one.
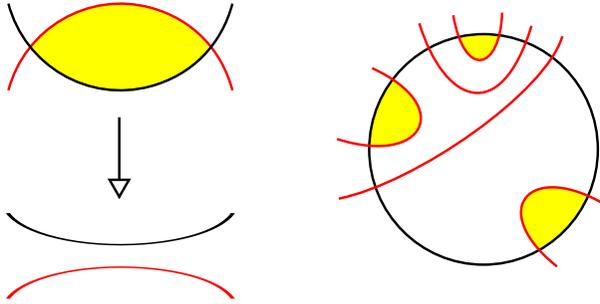
Figure 6.19. A bigon can be eliminated via homotopies (left). If γ_1 bounds a disc and γ_2 intersects γ_1, there is a bigon (right).

closed curves in S_g are in minimal position? We now prove a nice and simple criterion: they are in minimal position if and only if they do not form *bigons*.

The complement of two transverse simple closed curves is a finite disjoint union of open sets with polygonal boundaries; one such set is a *bigon* if it is a disc with two sides as in Figure 6.18.

Theorem 6.3.10 (Bigon criterion). *Two transverse simple closed γ_1, γ_2 in S_g are in minimal position if and only if they do not form bigons.*

Proof. If γ_1 and γ_2 create a bigon, the homotopies in Figure 6.19-(left) transform γ_1 and γ_2 into two curves that intersect in a smaller number of points: hence γ_1 and γ_2 are not in minimal position.

Suppose now that γ_1 and γ_2 do not form bigons: we must show that they are in minimal position. If γ_1 is trivial, it bounds a disc D as in Figure 6.19-(right). If γ_2 intersects γ_1, an *innermost* argument shows that they form a bigon: the curve γ_2 intersects D in arcs, each dividing D into two parts; if one part contains no other arc it is a bigon, otherwise we iterate. Therefore γ_1 and γ_2 are disjoint and hence in minimal position.

It remains to consider the case where both γ_1 and γ_2 are non-trivial. The torus case $g = 1$ is obtained by readapting the proof of Proposition 6.3.2 and is left as an exercise, so we suppose $g \geqslant 2$.

Figure 6.20. The lifts of γ_i and $\bar{\gamma}_i$ have distance bounded
by R and hence have the same endpoints (left). Two curves
that intersect in more than one point form a bigon (right).

Fix an arbitrary hyperbolic metric $S_g = \mathbb{H}^2/_\Gamma$ and let $\pi\colon \mathbb{H}^2 \to \mathbb{H}^2/_\Gamma$
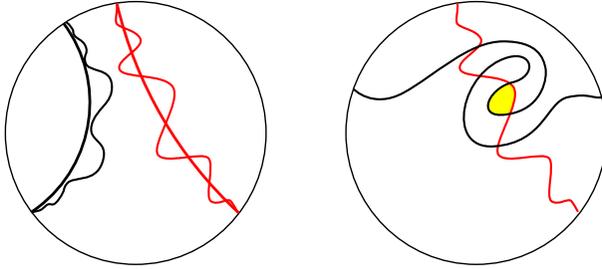be the projection. The closed curves γ_1 and γ_2 are now homotopic to two
simple closed geodesics $\bar{\gamma}_1$ and $\bar{\gamma}_2$. The lifts of γ_i and $\bar{\gamma}_i$ in $\mathbb{H}^2$ are *arcs* and
lines and there is a $R > 0$ such that every arc lies in the R-neighbourhood
of a line, see the proof of Proposition 6.3.8. Arcs and lines have the same
endpoints at infinity as in Figure 6.20-(left).

Two distinct arcs may intersect at most in one point: if they intersect
more, an innermost argument shows that they form a bigon D as in Figure
6.20-(right), which projects to a bigon $\pi(D)$ in S_g. (It is not immediate
that $\pi(D)$ is a bigon! The two vertices of $\pi(D)$ might coincide, but this
is easily excluded because S_g is orientable.)

We now show how to count the intersections between γ_1 and γ_2 di-
rectly in $\mathbb{H}^2$. Let $C(\gamma_i) \subset \Gamma$ be the conjugacy class of all the hyperbolic
transformations corresponding to γ_i. We know that the lifts of $\bar{\gamma}_i$ are
precisely the axis of the hyperbolic transformations in $C(\gamma_i)$.

By Corollary 4.2.3 any two such axis are either incident or ultra-
parallel. Hence two lifts of γ_1 and γ_2 intersect (in a single point) if and only
if the corresponding lifts of $\bar{\gamma}_1$ and $\bar{\gamma}_2$ intersect (in a single point), and this
holds if and only if the endpoints are linked in $\partial\mathbb{H}^2$. We have established
two bijective correspondences

$$\pi^{-1}(\gamma_1) \cap \pi^{-1}(\gamma_2) \longleftrightarrow \pi^{-1}(\bar{\gamma}_1) \cap \pi^{-1}(\bar{\gamma}_2) \longleftrightarrow X$$

with

$$X = \big\{ (\varphi_1, \varphi_2) \in C(\gamma_1) \times C(\gamma_2) \;\big|\; \text{Fix}(\varphi_1) \text{ and Fix}(\varphi_2) \text{ are linked} \big\}.$$

The bijective correspondences are Γ-equivariant. We quotient by Γ and
find

$$\gamma_1 \cap \gamma_2 \longleftrightarrow \bar{\gamma}_1 \cap \bar{\gamma}_2 \longleftrightarrow X/_\Gamma.$$

The cardinality k of $X/_\Gamma$ depends only on the homotopy type of γ_1 and γ_2. Therefore any two curves homotopic to γ_1 and γ_2 will have at least these k intersections. Hence γ_1 and γ_2 are in minimal position.   □

The bigon criterion furnishes an efficient algorithm to calculate the geometric intersection of any pair γ_1 and γ_2 of simple closed curves in S_g: we put them in transverse position, and then we simplify bigons as much as we can. After finitely many steps we get two curves in minimal position.
    Geodesic representatives are typically efficient: they minimise lengths (see Proposition 4.1.15), and they also minimise mutual intersections.

Corollary 6.3.11. *Let $g \geqslant 2$ and S_g have a hyperbolic metric. Two simple closed geodesics with distinct supports are in minimal position.*

Proof. Two geodesics do not form bigons: if they do, the bigon lifts to a bigon between two lines in $\mathbb{H}^2$, but lines intersect at most once.   □

Here is a very simple application of the bigon criterion.

Corollary 6.3.12. *If two simple closed curves γ, η intersect transversely in one point, we have $i(\gamma, \eta) = 1$. In particular, they are both non-trivial.*

Proof. The curves γ and η cannot form bigons.   □

The following exercise shows that the geometric intersection distinguishes a trivial curve from a non-trivial one.

Exercise 6.3.13. If a simple closed curve γ is not trivial, there is another simple closed curve η such that $i(\gamma, \eta) > 0$.

Hint. Use Proposition 6.3.3 to transform γ into a comfortable curve and draw a η which intersects γ in at most 2 points without bigons.   □

On the contrary, the algebraic intersection does *not* distinguish the trivial curve from any other separating curve.

Exercise 6.3.14. Let γ and η be non-trivial simple closed curves in the torus T of type (p, q) and (r, s). We have

$$i(\gamma, \eta) = \left| \det \begin{pmatrix} p & r \\ q & s \end{pmatrix} \right|.$$

6.3.8. Homotopy and isotopy of curves. We are now going to prove some "homotopy implies isotopy" theorems: we start with closed curves and we end in the next section with self-diffeomorphisms of surfaces. These slightly technical theorems have the remarkable and pleasant effect of making life easier in dimensions two and three (their impact to three-dimensional topology will be clear in the next chapter): topologists use these theorems everyday when they manipulate two- and three-manifolds.
    We start by showing that two non-trivial simple closed curves are homotopic if and only if they are isotopic. We consider a particular case.

Lemma 6.3.15. *Let γ_1 and γ_2 be two non-trivial simple closed curves in S_g. If they are disjoint and homotopic, they are parallel.*

Proof. Cut S_g along $\gamma_1 \cup \gamma_2$. We do not obtain discs because the curves are non-trivial and if we obtain an annulus the two curves are parallel. In all the other cases we obtain surfaces of negative Euler characteristic, and we may assign some hyperbolic metrics to them that glue to a hyperbolic metric on S_g where both γ_1 and γ_2 are distinct geodesics: a contradiction. $\square$

We now turn to the general case.

Proposition 6.3.16 (Homotopy implies isotopy). *Two non-trivial simple closed curves in S_g are homotopic if and only if they are isotopic.*

Proof. Let γ_1 and γ_2 be two non-trivial simple closed curves. With a small isotopy we put them in transverse position. Since $i(\gamma_1, \gamma_2) = i(\gamma_1, \gamma_1) = 0$ the two curves are either disjoint or form some bigon. If they form a bigon, we can eliminate it via isotopies as in Figure 6.19-(left) and after finitely many steps we get two disjoint curves.

The curves γ_1 and γ_2 are parallel by Lemma 6.3.15, and we use the annulus they cobound to move γ_2 isotopically over γ_1. The two curves now have the same support and the same orientation by Proposition 6.3.6: by Lemma 6.1.1 they are isotopic. $\square$

Warning 6.3.17. Two homotopic simple closed curves in a *three-manifold* are not necessarily isotopic, because they may be knotted differently: the *knot theory* studies precisely this phenomenon.

Corollary 6.3.18. *Let $g \geqslant 2$ and S_g have a hyperbolic metric. Every non-trivial simple closed curve in S_g is isotopic to a simple closed geodesic.*

6.3.9. Multicurves. Our next goal is to prove a "homotopy implies isotopy" theorem for self-diffeomorphisms of surfaces. The core of the proof is contained in the next section, and we will need there the notion of *multicurve*: we now introduce this simple concept and extend some of the previous results from curves to multicurves.

Definition 6.3.19. A *multicurve* μ in S_g is a finite set of disjoint non-trivial simple closed curves.

See an example in Figure 6.21, and note that every component of a multicurve is oriented. A multicurve is *essential* if it has no parallel components. By cutting S_g along an essential multicurve μ we get finitely many surfaces of negative Euler characteristic: if each such surface is a pair-of-pants, then μ is a *pants decomposition*, already considered in Section 6.2.2.
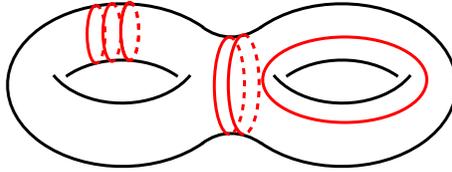
Figure 6.21. A multicurve in a surface of genus two.

Proposition 6.3.20. *An essential multicurve μ in S_g with $g \geqslant 2$ has at most $3g - 3$ components, and it has $3g - 3$ if and only if it is a pants decomposition.*

Proof. By cutting S_g along μ we get some surfaces $S^1, \ldots, S^k$ of negative Euler characteristic such that $\chi(S_g) = \chi(S^1) + \ldots + \chi(S^k)$. If each S^i is a pair-of-pants then $\chi(S^i) = -1$ and $k = -\chi(S_g) = 2g - 2$; the curves are $\frac{3}{2}(2g - 2) = 3g - 3$ because each curve separates two pants. If some S^i is not a pair-of-pants it can be further subdivided into pair-of-pants. $\square$

We define the *geometric intersection* $i(\mu_1, \mu_2)$ of two multicurves μ_1 and μ_2 as the minimum number of intersections of two transverse multicurves μ_1', μ_2' isotopic to μ_1, μ_2. This definition extends the geometric intersection of simple closed curves (the original definition for simple closed curves is with "homotopic" instead of "isotopic", but these are equivalent by Proposition 6.3.16).

Two transverse multicurves μ_1 and μ_2 are in *minimal position* if they intersect exactly in $i(\mu_1, \mu_2)$ points: the bigon criterion easily extends to this context.

Proposition 6.3.21. *Let $\mu_1, \mu_2 \subset S_g$ be two transverse multicurves. The following equality holds:*

$$i(\mu_1, \mu_2) = \sum_{\substack{\gamma_1 \subset \mu_1 \\ \gamma_2 \subset \mu_2}} i(\gamma_1, \gamma_2)$$

where the sum is taken on all components γ_1, γ_2 of μ_1, μ_2. The multicurves μ_1 and μ_2 are in minimal position if and only if they do not form bigons.

Proof. If μ_1 and μ_2 form no bigons, then any two components γ_1 and γ_2 form no bigons too (exercise) and are therefore in minimal position. This proves the equality and that μ_1 and μ_2 are in minimal position. $\square$

The formula says that if we consider a multicurve as a "sum" of disjoint simple closed curves, then the geometric intersection i behaves like a bilinear form: this viewpoint will be further explored in Chapter 8.
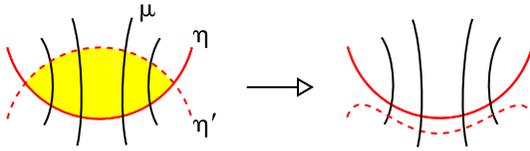
Figure 6.22. A bigon between η and η' intersects μ in vertical arcs and can be removed via an ambient isotopy that preserves the support of μ.

Note again that $i(\mu, \mu) = 0$. We extend Proposition 6.3.16 to essential multicurves.

Proposition 6.3.22 (Homotopy implies isotopy). *Let*

$$\mu_1 = \{\gamma_{1,1}, \ldots, \gamma_{1,n}\}, \quad \mu_2 = \{\gamma_{2,1}, \ldots, \gamma_{2,n}\}$$

be two essential multicurves in S_g. If $\gamma_{1,i}$ is homotopic to $\gamma_{2,i}$ for all i then there is an isotopy moving μ_1 to μ_2.

Proof. We adapt the proof of Proposition 6.3.16. Since $i(\gamma_{1,i}, \gamma_{2,j}) = i(\gamma_{1,i}, \gamma_{1,j}) = 0$ we get $i(\mu_1, \mu_2) = 0$ by Proposition 6.3.21 and after an isotopy that destroys the bigons we get $\mu_1 \cap \mu_2 = \varnothing$. Then Lemma 6.3.15 implies that $\gamma_{1,i}$ and $\gamma_{2,i}$ are parallel and we are done.   □

Corollary 6.3.23. *Let $g \geqslant 2$ and S_g have a hyperbolic metric. Every essential multicurve can be isotoped to a (unique) geodesic essential multicurve.*

In particular every pants decomposition straightens to a (unique) geodesic one: this fact will be used in Chapter 7 to parametrize all the hyperbolic metrics on a given S_g.

6.3.10. Minimal position. We end this discussion on multicurves by showing that the minimal position is unique up to isotopy.

Proposition 6.3.24. *Let μ, η be two multicurves in minimal position in S_g. The union $\mu \cup \eta$ of their supports depends up to ambient isotopy only on the isotopy classes of μ and η.*

Proof. Let μ', η' be multicurves in minimal position, individually isotopic to μ, η. We need to prove that the supports $\mu' \cup \eta'$ and $\mu \cup \eta$ are ambiently isotopic.

By hypothesis there is an isotopy carrying μ' to μ, which is ambient since multicurves are compact: hence we can suppose $\mu' = \mu$. We now construct an isotopy that fixes $\mu' = \mu$ as a set (not pointwise!) and carries η' to η.

Up to a small ambient isotopy fixing the set μ we may suppose that η and η' intersect transversely. If $\eta \cap \eta' \neq \varnothing$ then η and η' form a bigon

as in Figure 6.22-(left): the multicurve μ intersects the bigon in arcs that join distinct edges as in the figure (because $\mu' = \mu$ forms no bigons with η and η'). We can eliminate the bigon by an ambient isotopy that fixes μ as a set, as shown in Figure 6.22.

After finitely many steps we get $\eta \cap \eta' = \varnothing$. Since η and η' are disjoint and isotopic, every component of η is parallel to a component of η'. A maximal set of parallel curves in $\eta \cup \eta'$ is contained in a bigger annulus $[-1, 1] \times S^1$ which intersects μ either into arcs $[-1, 1] \times \{pt\}$ or into circles $\{pt\} \times S^1$. In both cases we easily see that $\mu \cup \eta$ is ambiently isotopic to $\mu \cup \eta'$. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad$ $\square$

As an example, consider two essential multicurves η and μ in S_g with $g \geqslant 2$ (for instance, two non-trivial simple closed curves). A hyperbolic metric on S_g produces two geodesic representatives $\bar{\eta}$ and $\bar{\mu}$ for them, and the following holds:

Corollary 6.3.25. *The support $\bar{\eta} \cup \bar{\mu}$ in S_g does not depend (up to ambient isotopy) on the chosen hyperbolic metric.*

Proof. With the exception of the common components of $\bar{\eta}$ and $\bar{\mu}$, the rest intersects in minimal position. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad$ $\square$

This corollary is one of the many instances of the following nice phenomenon: a hyperbolic metric on S_g may be used as an auxiliary structure to define objects or to prove statements which turn out *a posteriori* not to depend on the chosen metric. We will employ this strategy many times in Chapter 8.

6.4. Homotopy and isotopy

We have proved that "homotopy implies isotopy" for non-trivial simple closed curves in S_g, and now we want to prove an analogous result for diffeomorphisms of S_g. A crucial ingredient is a theorem proved by Smale in the 1950s, which considers the self-diffeomorphisms of the disc D^2.

Theorem 6.4.1. *Two diffeomorphisms $\varphi, \psi : D^2 \to D^2$ that coincide on ∂D^2 are isotopic, via an isotopy that fixes ∂D^2 pointwise.*

There is a proof of Theorem 6.4.1 in the topological category that is surprisingly easy and works in any dimension.

6.4.1. The Alexander trick. The proof of the following fact is so immediate, that it has a name: it is usually called the *Alexander trick*.

Proposition 6.4.2 (Alexander trick). *Two homeomorphisms $\varphi, \psi : D^n \to D^n$ that coincide on ∂D^n are continuously isotopic, via an isotopy that fixes ∂D^n pointwise.*

Proof. We consider $f = \varphi \circ \psi^{-1}$ and id_{D^n} and define an isotopy that transforms f into id_{D^n} fixing ∂D^n. The following function does the job:

$$F(x, t) = \begin{cases} x & \text{if } \|x\| \geqslant t, \\ tf\left(\frac{x}{t}\right) & \text{if } \|x\| \leqslant t. \end{cases}$$

The proof is complete. $\qquad\square$

Unfortunately this proof does not work in the smooth category, because the function F is not smooth (and not smoothable) at $(0,0)$. Of course, as every continuous map, our F may be approximated by smooth functions, but the injectivity of the slices $F(\cdot, t)$ can be lost in the approximation.

Moreover, the statement in the smooth category is just false in dimension $n \geqslant 7$: there are self-diffeomorphisms of D^n that are *not* smoothly isotopic to id_{D^n} in high dimension, and this is connected to the existence of *exotic spheres*, differentiable manifolds that are homeomorphic but not diffeomorphic to S^n.

6.4.2. Self-diffeomorphisms of the disc. After this very brief excursion in the topological world, we turn back to the smooth category: every map considered so far is smooth by assumption like in the rest of the book.

We prove here Theorem 6.4.1. We need a fact on planar fields.

Proposition 6.4.3. *On a nowhere-vanishing vector field on $\mathbb{R}^2$, a maximal integral curve is never trapped in a compact set.*

Proof. Suppose by contradiction that a maximal integral curve $\alpha : I \to \mathbb{R}^2$ lies in a compact set K for all $t > t_0$. Its ω-*limit* $\omega(\alpha) \subset K$ is the set of all points $x \in K$ to which γ accumulates, that is the intersection of the closures of $\alpha(t_0, \infty)$ for all $t_0 \in I$. The ω-limit is the intersection of a filtration of compact sets and is hence non-empty. Pick a point $p \in \omega(\alpha)$.

Up to a local diffeomorphism, the vector field near p is constant vertical as in Figure 6.23-(1). The integral curves near p are vertical. Since $p \in \omega(\alpha)$, the curve α contains infinitely many of them that tend to the one containing p. Two subsequent ones determine a closed curve as in Figure 6.23-(2), that bounds a disc by the Jordan curve theorem.

We prove that such a disc cannot exist. If it existed, we could reverse all arrows and rotate it to obtain another disc as in Figure 6.23-(3), and by gluing the two discs we would construct as in Figure 6.23-(4) a nowhere-vanishing vector field on the sphere. Such a field of course does not exist since the sphere has non-zero Euler characteristic. $\qquad\square$

We can now prove Theorem 6.4.1.

Proof of Theorem 6.4.1. After composing with ψ^{-1} we can suppose that $\psi = \mathrm{id}$. Now $\varphi|_{S^1}$ is the identity. Every self-diffeomorphism of any compact manifold that fixes the boundary is isotopic to one which is the
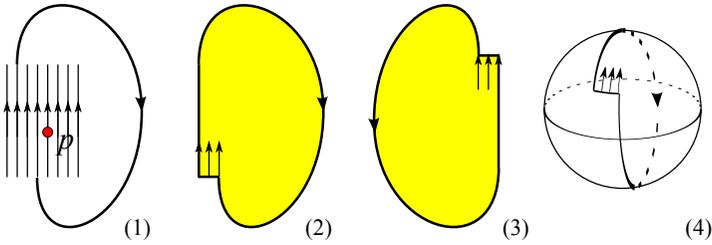
Figure 6.23. At a point p, after a local diffeomorphism we may suppose that the vector field is constantly vertical (1). If p is a limit point of an integral curve α we may find a closed curve that bounds a disc (2). If such a disc existed, we could rotate it and inverse all arrows (3) and glue the two portions to form a nowhere-vanishing vector field on the sphere (4).
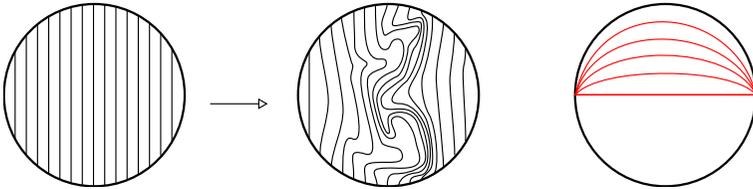
Figure 6.24. A self-diffeomorphism φ of D^2 fixing ∂D^2 pointwise (left). To get an isotopy fixing ∂D^2 pointwise we need to rescale via a diffeomorphism that maps each arc as in the right figure to itself, and varies smoothly from id to f_t^{-1} from bottom to top (right).

identity on a collar of the boundary: hence we can suppose that $\varphi = $ id on a collar of S^1.

We sketch φ in Figure 6.24-(left): vertical lines are transformed into paths with the same endpoints. We want to construct an isotopy between φ and id_{D^2} that "straightens" smoothly these lines, and we start by straightening their tangent vector fields.

Consider the constant unitary vertical vector field $X_0 = (0, 1)$ on D^2 and let X_1 be its image along the diffeomorphism φ. It is easy to construct a homotopy X_t between X_0 and X_1 through non-vanishing vector fields: we see both X_0 and X_1 as maps $D^2 \to \mathbb{C}^*$, we lift them to maps $D^2 \to \mathbb{C}$ via the universal cover $\exp \colon \mathbb{C} \to \mathbb{C}^*$, we make a convex combination and we project it back to $\mathbb{C}^*$. By construction the vectors X_t are constantly $(0, 1)$ on a collar of S^1.

We now carefully integrate the homotopy X_t of vector fields to an isotopy of diffeomorphisms. Let S_+^1 and S_-^1 be the upper and lower hemisphere of S^1. The vectors X_t at S_-^1 point inside and those at S_+^1 point

outside D^2. By Proposition 6.4.3 the integral curve γ_x^t of X_t starting from a point $(x, -\sqrt{1-x^2}) \in S_-^1$ exits at some point in S_+^1 after some time T_x^t that varies smoothly in x and t.

The point $\gamma_x^t(u)$ is defined for $u \in [0, T_x^t]$ and varies smoothly on x, t, u. We obviously have $T_x^0 = 2\sqrt{1-x^2}$ and also $T_x^1 = T_x^0$ since $\gamma_x^1 = \varphi \circ \gamma_x^0$. However T_x^t may not be constant in t, so we rescale the integral curves as

$$\tilde{\gamma}_x^t(u) = \gamma_x^t \left(u \cdot \frac{T_x^t}{2\sqrt{1-x^2}} \right).$$

The new curve $\tilde{\gamma}_x^t$ is defined in the interval $\left[0, 2\sqrt{1-x^2} \right]$ not depending on t. We can now define

$$\varphi_t\left(x, -\sqrt{1-x^2} + u\right) = \tilde{\gamma}_x^t(u).$$

The diffeomorphism $\varphi_t \colon D^2 \to D^2$ furnishes an isotopy between $\mathrm{id} = \varphi_0$ and $\varphi = \varphi_1$. However, the diffeomorphism φ_t does not fix the points in S_+^1 since the endpoint of $\tilde{\gamma}_x^t$ has a first coordinate $f_t(x)$ which might be distinct from x when $0 < t < 1$ (but varies smoothly on t and x).

To fix that it suffices to compose each φ_t with a diffeomorphism of D^2 that maps each arc as in Figure 6.24-(right) to itself and varies smoothly from id to f_t^{-1} from bottom to top. $\square$

Remark 6.4.4. If $\varphi \colon D^2 \to D^2$ is the identity on a collar of S^1, we can easily transform the isotopy φ_t so that it is the identity on this collar for all t. This will be useful to glue isotopies.

6.4.3. Homotopy and isotopy of diffeomorphisms. We are now ready to promote homotopies to isotopies of diffeomorphisms. We start with the easier sphere case, which is a simple consequence of Theorem 6.4.1.

Theorem 6.4.5. *Two diffeomorphisms $\varphi, \psi \colon S^2 \to S^2$ are isotopic $\iff$ they are homotopic $\iff$ they are co-oriented.*

Proof. We suppose that φ and ψ are co-oriented and we must prove that they are isotopic. Pick any disc $D \subset S^2$. Both φ and ψ send D to some disc in S^2. Theorem 1.1.14 furnishes an ambient isotopy relating these maps, so we may suppose that φ and ψ coincide on D. The closed complement is another disc and we conclude by Theorem 6.4.1.

More precisely, we pick a smaller closed disc $D' \subset \mathrm{int}(D)$ and consider the closed complement $D'' = S^2 \setminus \mathrm{int}(D')$. The diffeomorphisms φ and ψ of D'' coincide on a collar of $\partial D''$ and Theorem 6.4.1 together with Remark 6.4.4 furnishes an isotopy that is the identity on that collar. We extend this isotopy as the identity on $D \setminus D''$ and we are done. $\square$

The theorem has a very important three-dimensional application.

Corollary 6.4.6 (Smale's Theorem). *Every self-diffeomorphism of S^2 extends to a self-diffeomorphism of D^3.*
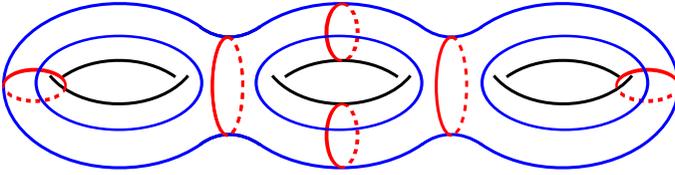
Figure 6.25. Two essential multicurves (blue and red) in minimal position, which subdivide the surface into hexagons.

Proof. Every self-diffeomorphism of S^2 is isotopic either to the identity or to a reflection by Theorem 6.4.5. We map a collar of S^2 to itself using this isotopy, then extend to the rest of D^3 using the identity or the reflection. $\qquad\square$

We can finally promote homotopies to isotopies on a general genus-g surface S_g. The proof uses many of the techniques introduced in the last sections: we employ multicurves to cut the surface into discs, and then use Theorem 6.4.1 individually on each disc.

Theorem 6.4.7. *Two diffeomorphisms φ, $\psi\colon S_g \to S_g$ are isotopic if and only if they are homotopic.*

Proof. We have seen the case $g = 0$, so we suppose $g \geqslant 1$. By composing with ψ^{-1} we can suppose that $\psi = \mathrm{id}$.

Fix two essential multicurves μ and η as in Figure 6.25. The figure easily generalises to any genus g, and the reader may check that μ and η subdivide the surface into hexagons (or a single square when $g = 1$). The important point is that $S_g \setminus (\mu \cup \eta)$ consists of polygons, none of which is a bigon.

By hypothesis φ is homotopic to the identity, so the image multicurves $\varphi(\mu)$ and $\varphi(\eta)$ are curve by curve homotopic to μ and η. By Proposition 6.3.22 the multicurve $\varphi(\mu)$ is isotopic to μ, and $\varphi(\eta)$ is isotopic to η.

The curves μ and η are in minimal position because there are no bigons. The images $\varphi(\mu)$ and $\varphi(\eta)$ are also in minimal position because φ is a diffeomorphism. By Proposition 6.3.24 the supports $\mu \cup \eta$ and $\varphi(\mu \cup \eta)$ are ambiently isotopic, so we may suppose that they coincide.

The graph $\mu \cup \eta$ is made of vertices (the intersections $\mu \cap \eta$) and edges. The components of μ and η are pairwise non-homotopic simple closed curves, hence φ sends necessarily every component to itself, and it does so orientation-preservingly by Proposition 6.3.6. This implies easily that vertices and edges are sent to themselves by φ. Hence $\varphi = \mathrm{id}$ on vertices and after an isotopy we may suppose that $\varphi = \mathrm{id}$ on edges too.

After an isotopy we may also suppose that $\varphi = \mathrm{id}$ on a regular neighbourhood U of $\mu \cup \eta$, obtained by thickening the cellularisation of $\mu \cup \eta$ to a handle decomposition with 0- and 1-handles. The complement of U

consists of discs (the hexagons). Consider one such disc D, enlarged a bit so that $\partial D \subset \mathrm{int}(U)$. The map φ sends D to itself and is the identity on a collar of ∂D. By Theorem 6.4.1 and Remark 6.4.4 there is an isotopy connecting φ to id on every such disc D that fixes pointwise this collar, so we can extend it constantly on the rest of U and get a global isotopy on S_g connecting φ and id.   □

This theorem has important consequences in dimensions 2 and 3.

6.5. Mapping class group

We have just proved that two self-diffeomorphisms of S_g are homotopic if and only if they are isotopic. We have also seen that when $g = 0$ there are only two self-diffeomorphisms up to isotopy: the identity and a reflection. Do we get a more complicated picture when $g \geqslant 1$? Yes, we get an interesting group, called the *mapping class group* of S_g. The group is naturally defined on all finite-type surfaces $S_{g,b,p}$.

6.5.1. Definition. Recall that the finite-type orientable surface $S_{g,b,p}$ has genus g, it has b boundary components, and p punctures.

Definition 6.5.1. The *mapping class group* of $S_{g,b,p}$ is the group

$$\mathrm{MCG}(S_{g,b,p}) = \mathrm{Diffeo}^+(S_{g,b,p})/_\sim$$

where $\mathrm{Diffeo}^+(S_{g,b,p})$ indicates the group of all orientation-preserving diffeomorphisms $S_{g,b,p} \to S_{g,b,p}$ that fix pointwise the boundary and $\varphi \sim \psi$ if φ and ψ are connected by an isotopy that fixes the boundary pointwise at every level.

Example 6.5.2. The groups $\mathrm{MCG}(S^2)$ and $\mathrm{MCG}(D^2)$ are trivial by Theorems 6.4.1 and 6.4.5.

The group $\mathrm{MCG}(S_{g,b,p})$ acts on $H_1(S_{g,b,p}, \mathbb{Z})$ since homotopic functions induce the same maps in homology. We get a group homomorphism

$$\mathrm{MCG}(S_{g,b,p}) \longrightarrow \mathrm{Aut}^+\big(H_1(S_{g,b,p}, \mathbb{Z})\big) = \mathrm{Aut}^+(\mathbb{Z}^n) = \mathrm{SL}_n(\mathbb{Z})$$

with $n = 2g + \max\{b + p - 1, 0\}$. This homomorphism is neither injective nor surjective in general. Its kernel is called the *Torelli group* of $S_{g,b,p}$.

The mapping class group of a sphere is trivial, so we look at the torus.

6.5.2. The torus. The mapping class group of the torus is a familiar group of 2×2 matrices.

Proposition 6.5.3. *The Torelli group of the torus T is trivial and*

$$\mathrm{MCG}(T) \cong \mathrm{Aut}^+(H_1(T)) = \mathrm{SL}_2(\mathbb{Z}).$$

Proof. Fix a meridian m and longitude l of T. A diffeomorphism φ of T that acts trivially on $H_1(T) = \pi_1(T) = \mathbb{Z}^2$ sends m and l to two simple closed curves $\varphi(m)$ and $\varphi(l)$ homotopic and hence isotopic to m
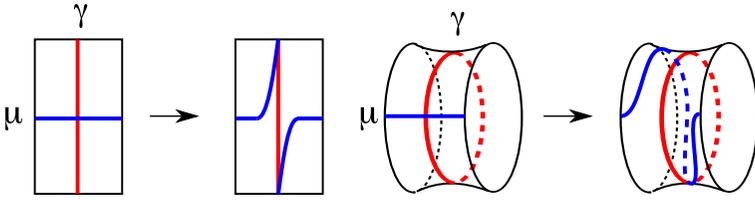
Figure 6.26. A Dehn twist along a curve γ maps a transverse arc μ onto an arc which makes a complete left turn.

and l: the proof of Theorem 6.4.7 shows that φ is isotopic to the identity. Therefore the Torelli group is trivial.

A matrix $A \in \mathrm{SL}_2(\mathbb{Z})$ acts linearly on $\mathbb{R}^2$ preserving the orientation and the lattice $\mathbb{Z}^2$ and hence descends to a self-diffeomorphism of $T = \mathbb{R}^2/\mathbb{Z}^2$. This shows that the map $\mathrm{MCG}(T) \to \mathrm{Aut}^+(H_1(T)) = \mathrm{SL}_2(\mathbb{Z})$ is surjective. $\qquad\square$

The mapping class group of a surface S_g of genus $g \geqslant 2$ is not isomorphic to a familiar group of matrices, at least as far as we know: indeed (except some low-genus cases) it is still unknown whether $\mathrm{MCG}(S_{g,b,p})$ is *linear*, *i.e.* isomorphic to a subgroup of $\mathrm{GL}(n, \mathbb{C})$ for some integer n.

For the moment we content ourselves with a concrete description of some particularly simple elements of $\mathrm{MCG}(S_{g,b,p})$ called *Dehn twists*.

6.5.3. Dehn twists. Let γ be a non-trivial simple closed curve in the interior of $S_{g,b,p}$. The *Dehn twist* along γ is the element $T_\gamma \in \mathrm{MCG}(S_{g,b,p})$ defined as follows.

Pick a tubular neighbourhood of γ orientation-preservingly diffeomorphic to $S^1 \times [-1, 1]$ where γ lies as $S^1 \times \{0\}$. Let $f : [-1, 1] \to \mathbb{R}$ be a smooth function which is zero in $\left[-1, -\frac{1}{2}\right]$ and 2π on $\left[\frac{1}{2}, 1\right]$. Let

$$T_\gamma : S_{g,b,p} \longrightarrow S_{g,b,p}$$

be the diffeomorphism that acts on the tubular neighbourhood as $T_\gamma(e^{i\alpha}, t) = (e^{i(\alpha + f(t))}, t)$ and on its complementary set in $S_{g,b,p}$ as the identity. We may visualise T_γ by noting that it gives a complete left turn to any arc μ that intersects γ as in Figure 6.26.

Proposition 6.5.4. *The element $T_\gamma \in \mathrm{MCG}(S_{g,b,p})$ is well-defined and depends only on the isotopy class of γ.*

Proof. In the definition of T_γ we have chosen a tubular neighbourhood for γ and a smooth function f. Tubular neighbourhoods are ambiently isotopic, and functions with fixed extremes are isotopic too: these facts imply easily that the isotopy class of T_γ is well-defined and depends only on the isotopy class of γ. $\qquad\square$

Remark 6.5.5. To define T_γ we needed the orientation of $S_{g,b,p}$, but not an orientation for γ. A bit surprisingly, if we change the orientation of γ the element T_γ remains unaffected.

The inverse T_γ^{-1} transforms every curve μ crossing γ via a complete *right*-turn and is sometimes called a *negative Dehn twist*.

We construct some examples in the torus T. Let m and l be some fixed meridian and longitude on the oriented T forming a positive basis, so that $m \cdot l = +1$ and we get an identification $\mathrm{MCG}(T) = \mathrm{SL}_2(\mathbb{Z})$.

Proposition 6.5.6. *The Dehn twists T_m and T_l are*

$$T_m = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}, \qquad T_l = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}.$$

Proof. In homology we find

$$T_m(m) = m, \quad T_m(l) = l - m, \qquad T_l(l) = l, \quad T_l(m) = m + l.$$

The proof is complete. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

By Exercise 6.3.4 there is a unique non-separating simple closed curve γ in S_g up to self-diffeomorphism for every $g \geq 1$. Therefore the Dehn twists T_γ along non-separating curves in S_g are all conjugate in $\mathrm{MCG}(S_g)$. In the torus case these can be easily identified algebraically.

Corollary 6.5.7. *An element $A \in \mathrm{SL}_2(\mathbb{Z}) = \mathrm{MCG}(T)$ is a (positive or negative) Dehn twist $\Longleftrightarrow$ it is primitive and $\mathrm{tr}A = 2$.*

Proof. Every such matrix in $\mathrm{SL}_2(\mathbb{Z})$ is conjugate to $\begin{pmatrix} 1 & \pm 1 \\ 0 & 1 \end{pmatrix}$. $\qquad\square$

Exercise 6.5.8. The Dehn twists T_m and T_l generate $\mathrm{MCG}(T)$.

We have seen that Dehn twists generate $\mathrm{MCG}(T)$, while many elements $A \in \mathrm{MCG}(T)$ are neither Dehn twists nor powers of Dehn twists (those with $\mathrm{tr}A \neq 2$). These two facts extend to higher-genus surfaces S_g; we start by generalising the first.

6.5.4. Dehn twists generate. Dehn twists are basic elements in the mapping class group, and we now prove that they generate the whole group. We restrict ourselves for simplicity to compact surfaces $S_{g,b}$.

Theorem 6.5.9. *The group $\mathrm{MCG}(S_{g,b})$ is generated by Dehn twists.*

To prove the theorem we will need some preliminary facts. We say that two non-separating simple closed curves in the interior of $S_{g,b}$ are *related* if there is a combination of isotopies and Dehn twists transforming the first into the second.

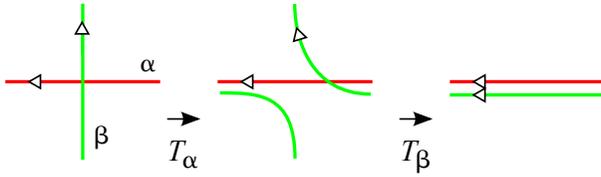Lemma 6.5.10. *The non-separating curves in $S_{g,b}$ are all related.*

Figure 6.27. Two curves α (red) and β (green) intersecting in one point are related: we get $\alpha = T_\beta(T_\alpha(\beta))$.
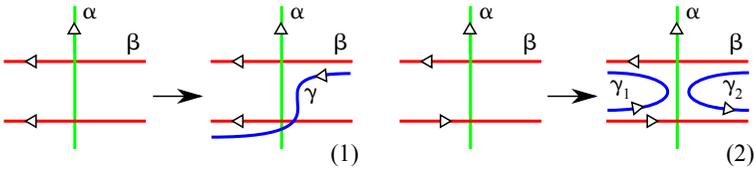

Figure 6.28. Pick two consecutive points in α (green) that intersect β (red). We can find a third non-separating curve γ (blue) intersecting α and β in $< k$ points. There are two cases (1) and (2): in (1) the curve γ intersects β in one point and is hence non-separating, in (2) we have two possibilities γ_1, γ_2, and one is certainly non-separating since β is.

Proof. Let α and β be two non-separating curves. Up to isotopy they intersect transversely into some k points. If $k = 1$ they are related by a couple of Dehn twists as shown in Figure 6.27. If $k = 0$, since they are both non-separating, one sees easily that there is another curve γ with $i(\alpha, \gamma) = i(\beta, \gamma) = 1$: hence α and β are both related to γ, so they are related themselves. If $k \geqslant 2$ we can find a curve γ intersecting α and β transversely into $< k$ points as shown in Figure 6.28 and we proceed by induction on k.   □

We now consider $S_{g,b}$ with $b \geqslant 2$ and fix two points p, q in two distinct boundary components of $S_{g,b}$. We consider all the properly embedded arcs in $S_{g,b}$ with endpoints at p and q. As above, we say that two such arcs are *related* if there is a combination of Dehn twists and isotopies transforming the first into the second.

Lemma 6.5.11. *The arcs in $S_{g,b}$ with endpoints at p and q are all related.*

Proof. Pick two arcs α and β and put them into transverse position: they intersect at their endpoints p and q and maybe transversely at some k other points. Figure 6.29 shows that after isotopies and Dehn twists we get $\alpha = \beta$.   □
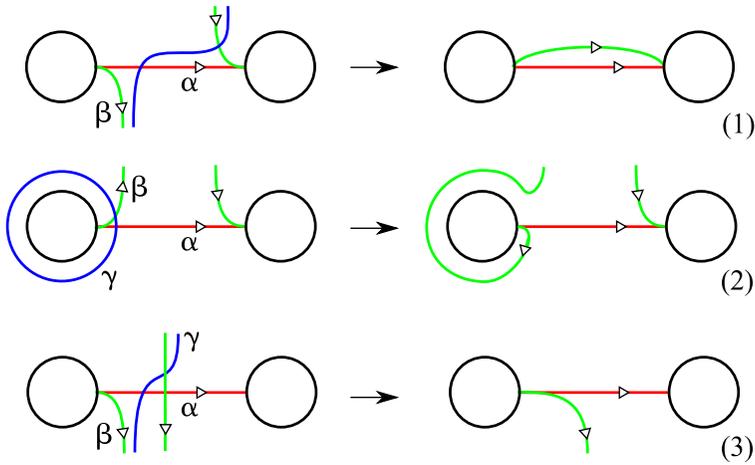
Figure 6.29. The arcs α (red) and β (green) intersect at their endpoints and maybe at some k interior points. If $k = 0$ and they are oriented as in (1), a Dehn twist along γ transforms β into α. If they are oriented as in (2), a negative Dehn twist along γ transforms this configuration into (1). If $k > 0$ we look at the first intersection point in α. If the orientations are coherent as in (3) a Dehn twist along γ decreases k. It the orientations are not coherent we change them as in (2).

We can finally prove Theorem 6.5.9.

Proof of Theorem 6.5.9. Let φ be a self-diffeomorphism of $S_{g,b}$ fixing pointwise the boundary. We prove that φ is generated by isotopies and Dehn twists.

We first prove the case $g = 0$ by induction on b. We know that $\mathrm{MCG}(S_{0,1})$ is trivial, so we suppose $b \geqslant 2$. Let p, q be points in distinct boundary components of $S_{0,b}$, and α be an arc connecting them. All the arcs with endpoints in p and q are related, and hence α and $\varphi(\alpha)$ are. Therefore up to composing with Dehn twists and isotopies we may suppose that φ is the identity of α and hence also on a tubular neighbourhood of α. We cut $S_{0,b}$ along α and get $S_{0,b-1}$, with φ transformed into a self-diffeomorphism of $S_{0,b-1}$. By induction on b the new φ is generated by Dehn twists and isotopies, so the original φ also is.

We prove the case $g > 0$ by induction on g. Let α be a non-separating simple closed curve. Since these are all related, up to isotopies and Dehn twists we may suppose that φ is the identity on α; as above we can cut $S_{g,b}$ along α, get $S_{g-1,b+2}$ and conclude by induction on g. $\qquad\square$

6.5.5. Action on simple closed curves. We now show that every element $\varphi \in \mathrm{MCG}(S_g)$ is determined by the way it permutes the (isotopy classes of) simple closed curves in S_g.

Proposition 6.5.12. *The action of $\varphi \in \mathrm{MCG}(S_g)$ on the isotopy classes of simple closed curves is faithful.*

Proof. The proof of Theorem 6.4.7 shows that if φ fixes the isotopy classes of two essential multicurves μ and η as in Figure 6.25 then it is isotopic to the identity. $\square$

Recall that every simple closed curve is oriented by assumption: if we considered *unoriented* simple closed curves the proposition would be false, because in low genus there are some φ that send every curve to its inverse: this happens for instance with the map $-I \in \mathrm{SL}_2(\mathbb{Z}) = \mathrm{MCG}(T)$.

6.5.6. References. The main source for this chapter is the book of Farb – Margalit [**18**], which contains a lot more information on the mapping class group of surfaces. We have also consulted Benedetti – Petronio [**4**] and Thurson's notes [**56**] and book [**57**], in particular for Smale's Theorem, that was originally proved in [**54**].

Teichmüller space

We have discovered in Chapter 6 that every surface S_g of genus $g \geqslant 2$ can be equipped with a hyperbolic metric, and we have already noticed that this metric is not unique: this chapter is entirely devoted to studying this non-uniqueness phenomenon.

It turns out that S_g admits a continuous family of non-equivalent hyperbolic metrics, that form altogether a nice topological space called the *Teichmüller space* of S_g. We prove in his chapter that the Teichmüller space of S_g is homeomorphic to an open ball of dimension $6g - 6$. To prove this fact we will introduce and study concepts like *length functions*, *earthquakes*, and *Fenchel–Nielsen coordinates*. The simple closed curves and their geodesic representatives will play a fundamental role in all the discussion.

7.1. Introduction

Let S_g be as usual a closed orientable surface of genus g. We know that S_g admits an elliptic, flat, or hyperbolic metric if and only if $g = 0$, $g = 1$, or $g \geqslant 2$ respectively. The elliptic metric on the two-sphere is unique up to isometries, but the flat and hyperbolic metrics on the other surfaces are not.

We want to define the space of all flat or hyperbolic metrics on S_g when $g \geqslant 1$. There are two natural ways do to this:

Definition 7.1.1. The *moduli space* of S_g is the set of all the flat or hyperbolic metrics on S_g considered up to orientation-preserving isometries and rescaling.

The *Teichmüller space* $\mathrm{Teich}(S_g)$ is the set of all the flat or hyperbolic metrics on S_g considered up to isometries isotopic to the identity and rescaling.

The *rescaling* of the metric is a simple operation that takes place only on the flat metrics on the torus T. On the torus T a flat metric g can be rescaled by any constant $\lambda > 0$ to give another flat metric λg: the rescaling changes the lengths by a factor $\sqrt{\lambda}$ and the area by a factor λ. Up to rescaling, we may for instance require that T has unit area.

At a first sight, the moduli space seems a more natural object to study. It turns out however that the Teichmüller space is homeomorphic
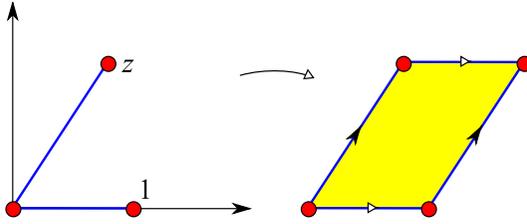
Figure 7.1. The flat metric on the torus T determined by $z \in H^2$ may be constructed by identifying the opposite sides of the parallelogram with vertices $0, 1, z, z + 1$. The lattice Γ is generated by 1 and z and the parallelogram is a fundamental domain.

(for some natural topology) to an open ball, while the moduli space is topologically more involved: it is then more comfortable to define and study the Teichmüller space first, and then consider the moduli space as a quotient of Teichmüller space.

7.1.1. The Teichmüller space of the torus. The flat metrics on the torus T are classified quite easily. We have seen in Proposition 3.4.12 that every flat torus T is of type $\mathbb{C}/\Gamma$ for some lattice $\Gamma < \mathbb{C}$ isomorphic to $\mathbb{Z}^2$.

Fix two generators m, l for $\pi_1(T)$. These are identified to two generators of Γ. Up to rescaling the metric, rotating $\mathbb{C}$ around the origin, and reflecting along the real axis, we may suppose that these generators are 1 and some complex number z lying in the upper half-plane H^2, so that $\Gamma = \langle 1, z \rangle$.

Proposition 7.1.2. *By sending the flat metric on T to z we get a bijection:*

$$\mathrm{Teich}(T) \longrightarrow H^2.$$

Proof. The map is well-defined: two metrics related by an isometry isotopic to the identity produce the same z. The inverse $H^2 \to \mathrm{Teich}(T)$ is constructed by identifying T with $\mathbb{C}/\langle 1, z \rangle$ sending (m, l) to $(1, z)$.   □

The flat metric that corresponds to $z \in H^2$ may be constructed by identifying the opposite sides of a parallelogram as in Figure 7.1. We will often tacitly identify $\mathrm{Teich}(T)$ with the projective plane H^2 via this bijective correspondence.

Remark 7.1.3. Let $T = \mathbb{C}/\Gamma$ be a flat torus. Every translation $z \mapsto z + w$ in $\mathbb{C}$ commutes with Γ and hence descends to an isometry of T. Therefore the isometry group $\mathrm{Isom}^+(T)$ is not discrete: every flat torus is *homogeneous*, *i.e.* for every pair of points $x, y \in T$ there is an isometry sending x to y.

Remark 7.1.4. Let $T = \mathbb{C}/_\Gamma$ be a flat torus. Every non-trivial element $\gamma \in \pi_1(T)$ is represented by a closed geodesic, unique up to translations. Its counterimage in $\mathbb{C}$ consists of parallel lines whose slope depends only on γ. The geodesic is simple if and only if γ is primitive.

7.1.2. Action of the mapping class group. Recall that the mapping class group $\mathrm{MCG}(S_g)$ of S_g is the group of all the orientation-preserving self-diffeomorphisms of S_g considered up to isotopy (or equivalently, homotopy). We now show that the mapping class group of S_g acts on its Teichmüller space.

A diffeomorphism $\varphi\colon S_g \to S_g$ transports a metric m on S_g into a new metric $\varphi_* m$ by pushing it forward as follows:

$$(\varphi_* m)_{\varphi(x)}\big(d\varphi_x(v), d\varphi_x(w)\big) = m_x(v, w).$$

If m varies through an isotopy, the metric $\varphi_* m$ varies through a corresponding isotopy: therefore φ acts on $\mathrm{Teich}(S_g)$ as follows

$$\mathrm{Teich}(S_g) \longrightarrow \mathrm{Teich}(S_g)$$
$$[m] \longmapsto [\varphi_* m]$$

If we vary φ by an isotopy the action is unaffected. Therefore the mapping class group $\mathrm{MCG}(S_g)$ acts on $\mathrm{Teich}(S_g)$, and by definition the quotient

$$\mathrm{Teich}(S_g)/_{\mathrm{MCG}(S_g)}$$

is the moduli space of S_g.

Everything can be written explicitly for the torus. Recall that we have identified $\mathrm{Teich}(T)$ with H^2 and $\mathrm{MCG}(T)$ with $\mathrm{SL}_2(\mathbb{Z})$, see Proposition 6.5.3.

Proposition 7.1.5. *The action of* $\mathrm{MCG}(T)$ *on* $\mathrm{Teich}(T)$ *is the following action of* $\mathrm{SL}_2(\mathbb{Z})$ *on* H^2 *as Möbius transformations:*

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} : z \longmapsto \frac{az - b}{-cz + d}.$$

Proof. The metric z assigns to T the structure $\mathbb{R}^2/_\Gamma$ with $\Gamma = \langle 1, z \rangle$ and (m, l) sent to $(1, z)$.

Pick $\varphi = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z}) = \mathrm{MCG}(T)$. Since $\varphi^{-1} = \left(\begin{smallmatrix} d & -b \\ -c & a \end{smallmatrix}\right)$, in the new metric $\varphi_*(z)$ we assign to (m, l) the translations $(d - cz, -b + az)$, which transform via rotations and dilations into $(1, \frac{az-b}{-cz+d})$. $\qquad\square$

We note in particular that $\mathrm{MCG}(T)$ acts via isometries on the hyperbolic plane H^2. The kernel of the action is $\{\pm I\}$: two matrices A and $-A$ act in the same way on $\mathrm{Teich}(T)$.

Corollary 7.1.6. *The moduli space of* T *is the orbifold* $H^2/_{\mathbb{P}\mathrm{SL}_2(\mathbb{Z})}$.

The orbifold $H^2/_{\mathbb{P}\mathrm{SL}_2(\mathbb{Z})}$ is described in Figure 3.22. It has two singular points of order 2 and 3: these represent the square and hexagonal torus, see Section 3.4.8 and Figure 3.15.
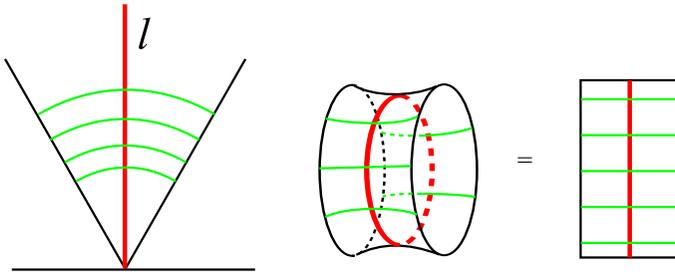
Figure 7.2. A R-annulus around a simple closed geodesic γ on a hyperbolic surface is the quotient of a R-neighbourhood of a line l by a hyperbolic transformation. The orthogonal (green) geodesic segments are parametrized by arc length as $[-R, R]$, hence the R-annulus is naturally parametrized as $S^1 \times [-R, R]$.

We have constructed a bijection between $\text{Teich}(T)$ and H^2, and we now want to construct for $g \geqslant 2$ an analogous identification between $\text{Teich}(S_g)$ and some open set of $\mathbb{R}^N$ for some N depending on g. To this purpose we need to introduce some concepts.

7.2. Earthquakes and length functions

Simple closed geodesics are a formidable tool to study $\text{Teich}(S_g)$: we can use a simple closed geodesic to *twist* a metric (the operation is called an *earthquake*), and by simply looking at the lengths of the other closed geodesics we can measure how the metric varies along this transformation. We introduce these operations here; later on, we will use them to parametrize $\text{Teich}(S_g)$.

7.2.1. Earthquakes. The hyperbolic, flat, and spherical metrics on surfaces may be twisted along simple closed geodesics: this operation is called an *earthquake*.

Let m be a complete hyperbolic, flat, or elliptic metric on an oriented surface S and γ be a simple closed geodesic in S. Recall that γ is a map $\gamma \colon S^1 \to S$. Fix an angle $\theta \in \mathbb{R}$. Informally, a new complete hyperbolic, flat, or elliptic metric m_θ on S is constructed by cutting S along γ and regluing with a counterclockwise twist of angle θ. Formally, the new metric is defined as follows.

In the hyperbolic case, recall from Proposition 4.1.16 that γ has a R-neighbourhood isometric to a R-tube for some $R > 0$. A R-tube in dimension two is a R-annulus as in Figure 7.2, defined by quotienting a R-neighbourhood of a line l by a hyperbolic transformation. The R-annulus is naturally parametrized as $S^1 \times [-R, R]$, where $\{e^{it}\} \times [-R, R]$ is the
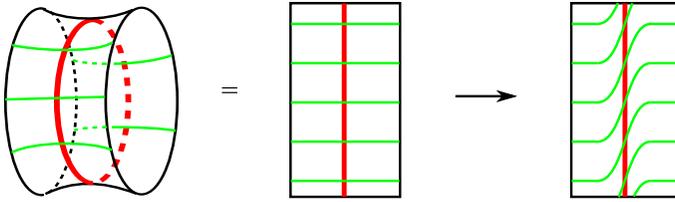
Figure 7.3. To define the earthquake we pick a diffeomorphism of the R-annulus that modifies the orthogonal segments as shown here.

geodesic segment orthogonal to γ in $\gamma(e^{it})$ parametrized by arc length. The flat and elliptic cases are analogous.

We choose a diffeomorphism φ of $S^1 \times [-R, R]$ that curves the segments left-wise with step θ as in Figure 7.3-(right). More precisely, let $f : [-R, R] \to \mathbb{R}$ be a smooth function which is zero on $\left[-R, -\frac{R}{2}\right]$ and is constantly θ on $\left[\frac{R}{2}, R\right]$. We set $\varphi(e^{it}, s) = (e^{i(t+f(s))}, s)$.

We define a new metric m_θ on S_g as follows: the metric tensor m_θ coincides with $\varphi_* m$ on the R-annulus and coincides with m on the complement of the $\frac{R}{2}$-annulus $S^1 \times \left[-\frac{R}{2}, \frac{R}{2}\right]$.

Proposition 7.2.1. *The metric tensor m_θ is well-defined and gives a complete hyperbolic, flat, or elliptic metric to S_g.*

Proof. It is well-defined because m and m_θ coincide on $S^1 \times \left[\frac{R}{2}, R\right]$, since $(e^{it}, s) \mapsto (e^{i(t+\theta)}, s)$ is an isometry of the R-annulus. It is hyperbolic, flat, or elliptic because both patches m and $\varphi_* m$ are. $\square$

Remark 7.2.2. In the new metric m_θ the curve γ is still a simple closed geodesic of the same length as before, and its R-neighbourhood is also unchanged.

Of course by deforming an elliptic metric in this way we get nothing new, because all the elliptic metrics on S^2 are isometric. Earthquakes are interesting only in the flat and hyperbolic geometries.

7.2.2. The earthquake map. We consider the surface S_g with $g \geqslant 1$ and show that earthquakes define nice actions on the Teichmüller space $\mathrm{Teich}(S_g)$.

If m is any flat or hyperbolic metric and γ is a non-trivial simple closed curve in S_g, we define m_θ^γ to be the flat or hyperbolic metric obtained from m via an earthquake of angle θ performed along the unique (up to translations if $g = 1$) simple closed geodesic homotopic to γ in the metric m.

Proposition 7.2.3. *The* earthquake map

$$E_\gamma \colon \mathbb{R} \times \mathsf{Teich}(S_g) \longrightarrow \mathsf{Teich}(S_g)$$
$$(\theta, m) \longmapsto m_\theta^\gamma$$

is a well-defined action of $\mathbb{R}$ on $\mathsf{Teich}(S_g)$. The map E_γ depends only on the homotopy class of γ.

Proof. The only ambiguity in the definition of m_θ^γ is the choice of the function $f \colon [-R, R] \to \mathbb{R}$. If we use another function f' the resulting metric $(m_\theta^\gamma)'$ changes only by an isotopy: the diffeomorphism of S_g which is the identity outside the R-annulus and sends (e^{it}, s) to $(e^{i(t+f(s)-f'(s))}, s)$ is an isometry between m_θ^γ and $(m_\theta^\gamma)'$, and is clearly isotopic to the identity.

To prove that E_γ is an action we need to check that

$$m_{\theta+\theta'}^\gamma = \left(m_{\theta'}^\gamma\right)_\theta^\gamma.$$

By Remark 7.2.2 we can take the same R-annulus to compose two earthquakes and the equality follows. ☐

Like the Dehn twists T_γ defined in Section 6.5.3, the action E_γ depends on the orientation of S_g but not on the orientation of γ. There is indeed a strong relation between earthquakes and Dehn twists on γ: as objects acting on Theichmüller space, the first generalise the second.

Proposition 7.2.4. *We have $T_\gamma(m) = E_\gamma(2\pi, m)$.*

Proof. It follows directly from the definitions. ☐

We now want to study the Teichmüller space and the action of E_γ on it. In mathematics a space is often beautifully described by some natural functions defined on it: this role is played here by the length functions of closed curves.

7.2.3. Length functions. A homotopically nontrivial (possibly non simple) closed curve γ in S_g defines a *length function*

$$\ell^\gamma \colon \mathsf{Teich}(S_g) \to \mathbb{R}_{>0}$$

which assigns to a metric $m \in \mathsf{Teich}(S_g)$ the length $\ell^\gamma(m)$ of the unique closed geodesic homotopic to γ.

When $g = 1$ we must actually specify a couple of things in the definition: the closed geodesic γ is unique only up to translations, which do not affect its length; on the other hand rescaling *does* affect lengths, so to get a well-defined length function we rescale m to have unit area.

We want to study these length functions, and as usual we first analyse the simpler torus flat world, where everything can be described explicitly.
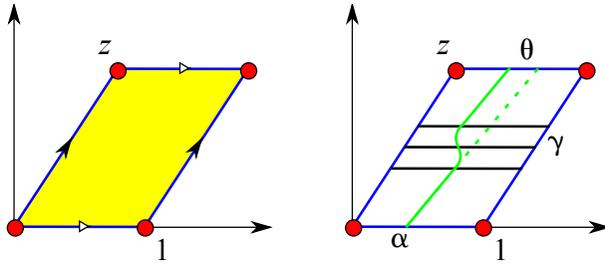
Figure 7.4. A torus with metric z (left) twisted along the horizontal curve γ (right). The curve α is a closed geodesic in the new metric.

7.2.4. Length functions on the torus. We denote every (isotopy class of) non-trivial simple closed curve on T with a coprime pair (p, q) of integers and we identify the Teichmüller space $\mathrm{Teich}(T)$ with $H^2 \subset \mathbb{C}$, see Propositions 6.3.2 and 7.1.2. The length functions may be written explicitly.

Proposition 7.2.5. *The formula holds:*

$$\ell^{(p,q)}(z) = \frac{|p + qz|}{\sqrt{\Im z}}$$

for every simple closed curve (p, q) and every metric $z \in H^2$.

Proof. Up to rescaling we have $T = \mathbb{R}^2/\Gamma$ with $\Gamma = \langle 1, z \rangle$. The translation in Γ corresponding to (p, q) is $p \cdot 1 + q \cdot z$ and the closed geodesic it produces has length $|p + qz|$. The area of the torus T is $\Im z$ (see a fundamental domain in Figure 7.1) and hence we must rescale it by $1/\sqrt{\Im z}$.   □

We can also write the earthquake action on the meridian $m = (1, 0)$.

Exercise 7.2.6. We have:

$$E_m(\theta, z) = z + \frac{\theta}{2\pi}.$$

Hint. Look at the closed geodesic α in Figure 7.4.   □

We have discovered in particular that E_m is a parabolic isometry of H^2, and we now consider the earthquake action along a generic curve (p, q).

Corollary 7.2.7. *The earthquake action $E_{(p,q)}$ is the 1-parameter family of parabolic transformations with fixed point $-\frac{p}{q} \in \partial H^2$.*

Proof. We know this when $(p, q) = (1, 0) = m$. In general, we send $(1, 0)$ to (p, q) via some element of the mapping class group: this element acts by isometries of H^2 by Proposition 7.1.5 and sends ∞ to $-\frac{p}{q}$.   □

The orbits of $E_{(p,q)}$ are the horospheres centred at $-\frac{p}{q}$. We have discovered that, quite unexpectedly, the hyperbolic geometry of the plane H^2 is well designed to model the Teichmüller space of the flat torus T.

It is now natural to identify the *unoriented* simple closed curve $\pm(p,q)$ with the rational point $-\frac{p}{q}$ in $\mathbb{R} \cup \{\infty\} \subset \partial H^2$. We get the following convexity property.

Corollary 7.2.8. *Let γ, η be two simple closed curves on T. If $i(\gamma, \eta) > 0$ the length function ℓ^η is strictly convex on the orbits of E_γ.*

Proof. We may suppose $\gamma = m = (1,0)$ and note that the condition $i(\gamma, \eta) > 0$ translates into $\eta = (p,q) \neq (\pm 1, 0)$. The function in Proposition 7.2.5 is strictly convex on the horospheres $\Im z = k$ when $(p,q) \neq (\pm 1, 0)$. $\qquad\square$

Summing up, the torus picture is the following:
- the mapping class group acts on the Teichmüller space roughly like $\mathbb{PSL}_2(\mathbb{Z})$ acts isometrically on the hyperbolic plane H^2,
- the unoriented simple closed curves are in 1-1 correspondence with the rational points in $\partial H^2 = \mathbb{R} \cup \infty$,
- earthquakes act like parabolic transformations centred at these rational points,
- the length function ℓ^γ is constant at the horospheres centred at γ but strictly convex at the horospheres centred at all the other curves.

The main goal of Chapters 7 and 8 is to draw a similar picture for surfaces S_g of higher genus $g \geqslant 2$. Everything is more difficult in the hyperbolic world, because there are no nice explicit formulas describing the length functions, the action of the mapping class group, the simple closed curves, and the earthquakes. We now start by generalising the last point: the strict convexity of length functions.

7.2.5. Convexity of the length functions. We consider S_g with $g \geqslant 2$. Our aim now is to generalise Corollary 7.2.8 to the hyperbolic setting and then later use this convexity property to parametrize $\mathrm{Teich}(S_g)$. We need a preliminary result.

Exercise 7.2.9. Let $f \colon \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}_{\geqslant 0}$ be strictly convex and proper. The function

$$F \colon \mathbb{R}^n \longrightarrow \mathbb{R}_{\geqslant 0}$$
$$y \longmapsto \min\left\{ f(x,y) \;\middle|\; x \in \mathbb{R}^m \right\}$$

is well-defined, strictly convex, and proper.

The following generalisation of Corollary 7.2.8 says that length functions are convex on the earthquakes orbits, and very often they are strictly convex and proper.
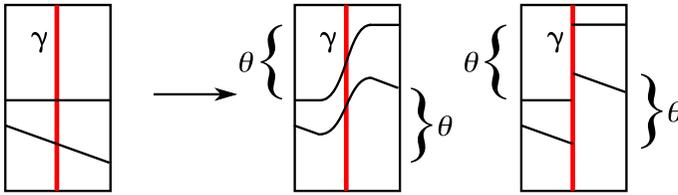
Figure 7.5. A geodesic in m_θ^γ can be seen on the original metric m as follows: it is a geodesic outside the annulus and deviates on the left by an angle θ every time it crosses it (centre). We may simplify the picture by describing it as a broken geodesic line that makes a left θ-jump each time that it crosses γ (right).

Proposition 7.2.10. *Let η and γ be two homotopically non-trivial simple closed curves in S_g and m be a hyperbolic metric on S_g. The function*

$$\mathbb{R} \longrightarrow \mathbb{R}_{\geqslant 0}$$
$$\theta \longmapsto \ell^\eta(m_\theta^\gamma)$$

is

- *constant if $i(\eta, \gamma) = 0$,*
- *strictly convex and proper if $i(\eta, \gamma) > 0$.*

Proof. The metric m_θ^γ is obtained by twisting m of an angle θ along the simple closed geodesic γ. If $i(\eta, \gamma) = 0$ the curves η and γ are disjoint geodesics in m and η is not affected by the earthquakes that we perform near γ, hence the function $\ell^\eta(m_\theta^\gamma)$ is constant.

Consider the case $n = i(\eta, \gamma) > 0$. Denote by $\bar\eta^\theta$ the geodesic representative of η in the twisted metric m_θ^γ: it intersects γ transversely in n points.

Fix a sufficiently small R-annulus around γ and note that the geodesics in m_θ^γ can be seen in the original metric m as follows: these are curves that are geodesic outside the R-annulus and deviate smoothly on the left each time they cross it as in Figure 7.5-(centre). We may substitute each smooth deviation with a broken jump as shown in Figure 7.5-(right) and get a bijection

$$\left\{ \begin{array}{c} \text{closed geodesics} \\ \text{with respect to } m_\theta^\gamma \end{array} \right\} \longleftrightarrow \left\{ \begin{array}{c} \text{broken geodesics} \\ \text{with respect to } m \end{array} \right\}$$

where a *broken geodesic* is a geodesic that at every crossing of γ jumps to the left at distance $\frac{\theta L(\gamma)}{2\pi}$ and then keeps going on, leaving γ with the same incidence angle (here $L(\gamma)$ is the length of γ). This correspondence is useful because it preserves the lengths: the length of the closed geodesic for m_θ^γ is equal to the length of the corresponding broken geodesic (which
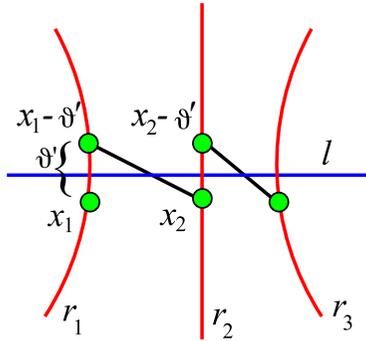
Figure 7.6. The (blue) line l is a lift of $\bar{\eta}$. Consider n consecutive intersections with lifts $r_1, \ldots, r_{n+1}$ of γ (in red), with $\tau(r_1) = r_{n+1}$, and parametrize each r_i with $\mathbb{R}$ via arc-length. We set $\theta' = \theta L(\gamma)/2pi$.

is the sum of the lengths of its components), because the segments in Figure 7.5-(left) and (right) are isometric.

We lift this description to the universal cover $\mathbb{H}^2$. We fix a lift l of $\bar{\eta} = \bar{\eta}^0$ and pick $n+1$ consecutive intersections $r_1, \ldots, r_{n+1}$ of l with the lifts of γ as in Figure 7.6. The hyperbolic transformation τ with axis l corresponding to η sends r_1 to r_{n+1}. We parametrise each r_i with $\mathbb{R}$ via arc length.

The closed geodesic $\bar{\eta}^\theta$, represented as a broken geodesic, lifts to a broken geodesic which starts at some point $x_1 - \theta L(\gamma)/2\pi \in r_1$ and arrives at some other point $x_2 \in r_2$, then jumps on the left at distance $\theta L(\gamma)/2\pi$ and starts again from $x_2 - \theta L(\gamma)/2\pi$, and so on until it reaches the point $\tau(x_1) \in \tau(r_1) = r_{n+1}$. If we make the points $x_1 \in r_1, \ldots, x_n \in r_n$ vary we get various broken paths in this way, but only one arrives and exits from each line r_i with the same incidence angles and thus represents $\bar{\eta}^\theta$. The other broken paths represent piecewise-geodesic curves homotopic to $\bar{\eta}^\theta$ and are therefore longer than $\bar{\eta}^\theta$. Hence

$$\ell^\eta(m_\theta^\gamma) = \min\left\{ \sum_{i=1}^n d\left(x_i - \frac{\theta L(\gamma)}{2\pi}, x_{i+1} \right) \;\middle|\; (x_1, \ldots, x_n) \in \mathbb{R}^n \right\}$$

where $x_{n+1} = \tau(x_1)$. We can now prove that the function $\theta \mapsto \ell^\eta(m_\theta^\gamma)$ is proper and strictly convex. The function

$$\psi \colon \mathbb{R}^{2n} \longrightarrow \mathbb{R}$$

$$(x_1, y_1, \ldots, x_n, y_n) \longmapsto \sum_{i=1}^n d(y_i, x_{i+1})$$

where $x_{n+1} = \tau(x_1)$ is strictly convex and proper by Proposition 2.4.4. The auxiliary function

$$\phi \colon \mathbb{R}^{2n} \times \mathbb{R} \longrightarrow \mathbb{R}$$
$$(x, \theta) \longmapsto \psi(x)$$

is only convex, but its restriction to the subspace

$$H = \left\{ y_i = x_i - \frac{\theta l(\gamma)}{2\pi} \right\}$$

is strictly convex and proper, because the subspace H is not parallel to the direction $(0, \ldots, 0, 1)$. The coordinates x_i and θ identify H with $\mathbb{R}^n \times \mathbb{R}$. The restriction $f = \phi|_H$ is hence a function $f \colon \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ and we obtain

$$\ell^\eta(m_\theta^\gamma) = \min \left\{ f(x, \theta) \mid x \in \mathbb{R}^n \right\}.$$

By Exercise 7.2.9 the function $\theta \mapsto \ell^\eta(m_\theta^\gamma)$ is strictly convex and proper. $\quad\square$

We now employ this convexity property to prove some facts on the earthquakes and the Teichmüller space.

7.2.6. Earthquakes on essential multicurves. We use the convexity of the length functions to prove the following. We suppose again that $g \geqslant 2$.

Corollary 7.2.11. *For every simple closed curve γ, the earthquake action E_γ on $\mathrm{Teich}(S_g)$ is free.*

Proof. Suppose by contradiction that $m = m_{\theta_0}^\gamma$ for some $\theta_0 > 0$. Then $m = m_{n\theta_0}^\gamma$ for every $n \in \mathbb{Z}$. Let η be a simple closed curve with $i(\eta, \gamma) > 0$, which exists by Exercise 6.3.13; the function $\theta \mapsto \ell^\eta(m_\theta^\gamma)$ is strictly convex and constant on $\{n\theta_0, n \in \mathbb{Z}\}$, a contradiction. $\quad\square$

The earthquake action is defined more generally for essential multi-curves. An essential multicurve $\mu = \gamma_1 \sqcup \cdots \sqcup \gamma_k$ of S_g determines an action

$$E_\mu \colon \mathbb{R}^k \times \mathrm{Teich}(S_g) \longrightarrow \mathrm{Teich}(S_g)$$
$$(\theta, m) \longmapsto m_\theta^\mu$$

where $\theta = (\theta_1, \ldots, \theta_k)$ and $m_\theta^\mu = m_{\theta_1}^{\gamma_1} \circ \cdots \circ m_{\theta_k}^{\gamma_k}$. Note that the actions on disjoint curves commute.

Corollary 7.2.12. *For every essential multicurve μ, the earthquake action E_μ on $\mathrm{Teich}(S_g)$ is free.*

Proof. We may complete μ to a pants-decomposition $\mu = \gamma_1 \sqcup \ldots \sqcup \gamma_{3g-3}$. Pick for every $i = 1, \ldots, 3g - 3$ a curve γ_i' as in Figure 7.7 such that $i(\gamma_i, \gamma_i') > 0$ for all i and $i(\gamma_i, \gamma_j') = 0$ for all $i \neq j$.

Suppose by contradiction that $m = m_\theta^\mu$ for some $\theta \neq 0$: hence $m_{n\theta}^\mu = m$ for all $n \in \mathbb{Z}$. There is an i such that $\theta_i \neq 0$. The length function
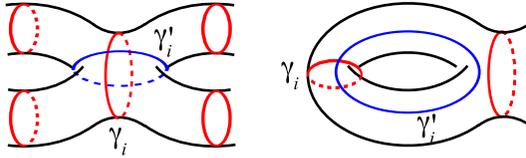
Figure 7.7. Choose for each component γ_i of a pants decomposition μ a curve γ_i' that intersects γ_i in one or two points and is disjoint from the other components of μ. There are two cases to consider, depending on whether the two pants adjacent to γ_i are distinct (left) or not (right).
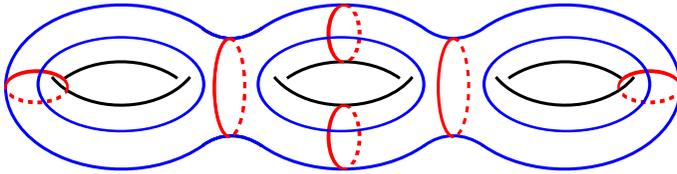
Figure 7.8. A frame for the Fenchel-Nielsen coordinates consists of a (red) pants decomposition μ and a (blue) transverse multicurve ν that cuts each pair of pants into two hexagons. The number of components of μ is $3g - 3$, that of ν can vary.

$\ell^{\gamma_i'}(m_\theta^\mu)$ depends only on θ_i and not on the other coordinates of θ: therefore it equals $\ell^{\gamma_i'}(m_{\theta_i}^{\gamma_i})$ which is strictly convex, a contradiction.   □

7.3. Fenchel–Nielsen coordinates

It is now time to fix a global set of coordinates for the Teichmüller space when $g \geqslant 2$. These are the *Fenchel–Nielsen coordinates* and they identify $\text{Teich}(S_g)$ with $\mathbb{R}^{6g-6}$, more precisely with $\mathbb{R}_{>0}^{3g-3} \times \mathbb{R}^{3g-3}$.

7.3.1. The coordinates. We want to construct a parametrisation for $\text{Teich}(S_g)$ when $g \geqslant 2$. To identify a finite-dimensional vector space with $\mathbb{R}^n$ one needs to fix a basis; likewise, here the parametrisation depends on the choice of a frame.

Let S_g be oriented. A *frame* for S_g consists of two essential multicurves μ and ν in minimal position, such that:

(1) the multicurve μ is a pants decomposition,
(2) the multicurve ν decomposes every pair-of-pants in two hexagons.

An example that generalises easily to any genus $g \geq 2$ is shown in Figure 7.8. The pants decomposition $\mu = \gamma_1 \sqcup \ldots \sqcup \gamma_{3g-3}$ consists of $3g - 3$ curves, while the number of curves in ν is not fixed a priori and
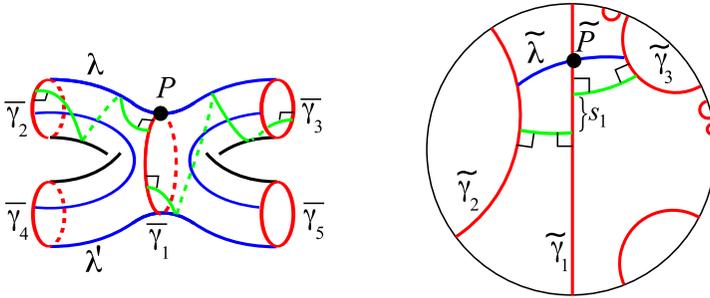
Figure 7.9. A closed geodesic $\bar{\gamma}_1$ and the two adjacent pairs-of-pants. The torsion parameter θ_1 measures the distance (in the universal covering) between two orthogeodesics (coloured in green) via the formula $\theta_i = \frac{2\pi s_i}{l_i}$.

depends on our choice of ν. We now show that a frame induces a *Fenchel–Nielsen* map

$$\text{FN}\colon \text{Teich}(S_g) \longrightarrow \mathbb{R}^{3g-3}_{>0} \times \mathbb{R}^{3g-3}$$

$$m \longmapsto (l_1, \ldots, l_{3g-3}, \theta_1, \ldots, \theta_{3g-3}).$$

The map FN is defined as follows. Let $m \in \text{Teich}(S_g)$ be a hyperbolic metric. The $3g - 3$ *length parameters* $l_i = \ell^{\gamma_i}(m)$ are defined using the length functions: the multicurve μ has a unique geodesic representative

$$\bar{\mu} = \bar{\gamma}_1 \sqcup \ldots \sqcup \bar{\gamma}_{3g-3}$$

in the metric m by Corollary 6.3.23, and l_i is the length of $\bar{\gamma}_i$. Note that these parameters depend only on μ and not on ν.

The *torsion angles* θ_i are more subtle to define: the geodesic multicurve $\bar{\mu}$ decomposes S_g into geodesic pairs-of-pants, and the angle θ_i measures somehow the way the two geodesic pairs-of-pants are glued along the closed geodesic $\bar{\gamma}_i$. The precise definition of θ_i needs the auxiliary multicurve ν.

We fix $i = 1$ for simplicity and define θ_1. Figure 7.9-(left) shows the two geodesic pants adjacent to $\bar{\gamma}_1$ (they might coincide). The second multicurve ν intersects these pants in four blue arcs, two of which λ, λ' intersect $\bar{\gamma}_1$: we pick one, say λ. We fix a lift $\tilde{P} \in \mathbb{H}^2$ of $P = \bar{\gamma}_1 \cap \lambda$ and we lift all the curves incident to P: the geodesic $\bar{\gamma}_1$ lifts to a line $\tilde{\gamma}_1$ and λ lifts to a (non-geodesic) curve $\tilde{\lambda}$ that connects two lifts $\tilde{\gamma}_2$ and $\tilde{\gamma}_3$ of the closed geodesics $\bar{\gamma}_2$ and $\bar{\gamma}_3$. See Figure 7.9-(right).

We draw as in the figure the unique orthogeodesics connecting $\tilde{\gamma}_1$ to $\tilde{\gamma}_2$ and $\tilde{\gamma}_3$ and we denote by s_1 the signed length of the segment in $\tilde{\gamma}_1$ comprised between these two orthogeodesics, with positive sign if (as in the figure) an observer walking on a orthogeodesic towards $\tilde{\gamma}_1$ sees the other orthogeodesic on its left (here we use the orientation of S_g).

Figure 7.10. If we pick λ' instead of λ we find a segment of the same length s_1. This holds because the two geodesic pairs-of-pants incident to γ_1 decompose into two hexagons isometric to A and B as shown, and the sides of A and B contained in $\bar{\gamma}_1$ have the same length $\frac{l(\gamma_1)}{2}$, that is half the length of γ_1.

By repeating this construction for each $\bar{\gamma}_i$ we find some real numbers s_i. Finally, the torsion parameter θ_i is

$$\theta_i = \frac{2\pi s_i}{l_i}.$$

Theorem 7.3.1 (Fenchel-Nielsen coordinates). *The map* FN *is well-defined and is a bijection.*

Proof. We first note that in the definition of the torsion parameters we could have chosen λ' instead of λ. We would have obtained the same length s_i as shown in Figure 7.10. Moreover a hyperbolic metric m' isometric to m through a diffeomorphism φ isotopic to the identity has the same parameters l_i and θ_j since they depend only on the isotopy classes of μ and ν. Therefore FN is well-defined.

We prove that FN is surjective. For every vector $(l_1, \ldots, l_{3g-3}) \in \mathbb{R}_{>0}^{3g-3}$ we may use Proposition 6.2.1 and construct a metric on S_g by assigning to each pair-of-pants of the pants decomposition μ the (unique) hyperbolic metric with boundary lengths l_i. We get a metric with some arbitrary torsion angles θ, which can be changed arbitrarily by an earthquake along μ: it is easy to check that an earthquake with angles θ' changes the torsion angles from θ to $\theta + \theta'$, hence any torsion parameter can be realised and FN is surjective.

We prove that FN is injective. If $FN(m) = FN(m')$, up to acting via earthquakes we suppose that $FN(m) = FN(m') = (l_1, \ldots, l_{3g-3}, 0, \ldots, 0)$. Since the torsion parameter is zero, the orthogeodesics in Figure 7.9-(right) match and project in S_g to a geodesic multicurve $\bar{\nu}$ isotopic to ν and
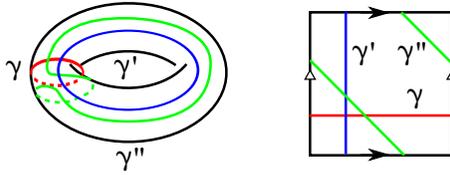
Figure 7.11. The curves γ (red), γ' (blue), and $\gamma'' = T_\gamma(\gamma')$ (green) on the torus.

orthogonal to $\bar{\mu}$. Therefore $S_g \setminus (\bar{\mu} \cup \bar{\nu})$ is a tessellation of S_g into right-angled hexagons, determined by the lengths l_i. Both metrics m and m' have the same tessellation and are hence isometric, via an isometry which is isotopic to the identity. $\square$

Remark 7.3.2. As shown in the proof, the torsion parameters for m are zero if and only if the geodesic representatives $\bar{\nu}$ and $\bar{\mu}$ of ν and μ are everywhere orthogonal.

7.3.2. Length functions of $9g - 9$ curves. It is now natural to ask whether the length functions determine every point in Teich(S_g). The answer is positive; as usual, to warm up we start by examining the torus.

Let γ, γ' be two simple closed curves in the torus T with $i(\gamma, \gamma') = 1$ and let $\gamma'' = T_\gamma(\gamma')$ be obtained by Dehn twisting γ' along γ, see Figure 7.11.

Proposition 7.3.3. *The map*

$$L : \text{Teich}(T) \longrightarrow \mathbb{R}^3_{>0}$$

$$m \longmapsto \left(\ell^\gamma(m), \ell^{\gamma'}(m), \ell^{\gamma''}(m) \right)$$

is injective.

Proof. After fixing γ and γ' as a homology basis, we have $\gamma = (1, 0)$, $\gamma' = (0, 1)$, and $\gamma'' = (1, -1)$. Proposition 7.2.5 gives

$$L(z) = \left(\frac{1}{\sqrt{\Im z}}, \frac{|z|}{\sqrt{\Im z}}, \frac{|z - 1|}{\sqrt{\Im z}} \right)$$

which is easily seen to be injective on Teich(T) $= H^2$. $\square$

A similar set of $9g - 9$ curves does the job on S_g when $g \geqslant 2$. Let $\mu = \gamma_1 \sqcup \ldots \sqcup \gamma_{3g-3}$ be a pants decomposition for S_g. For each γ_i we choose a curve γ_i' as in Figure 7.7, and we indicate by $\gamma_i'' = T_{\gamma_i}(\gamma_i')$ the curve obtained by Dehn-twisting γ_i' along γ_i, see an example in Figure 7.12.

Proposition 7.3.4. *The map*

$$L : \text{Teich}(S_g) \longrightarrow \mathbb{R}^{9g-9}_{>0}$$

$$m \longmapsto \left(\ell^{\gamma_i}(m), \ell^{\gamma_i'}(m), \ell^{\gamma_i''}(m) \right)$$
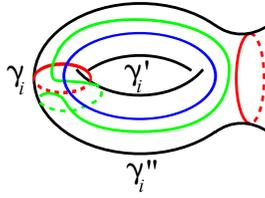
Figure 7.12. The curves γ_i (red), γ_i' (blue), and $\gamma_i'' = T_{\gamma_i}(\gamma_i')$ (green) when γ_i is adjacent twice to the same pair of pants. It is an instructive exercise to draw γ_i' and γ_i'' when γ_i is incident to distinct pair of pants.

is injective.

Proof. We compose L with FN^{-1} and obtain a map

$$L \circ \mathrm{FN}^{-1} \colon \mathbb{R}_{>0}^{3g-3} \times \mathbb{R}^{3g-3} \longrightarrow \mathbb{R}_{>0}^{9g-9}$$
$$(l_i, \theta_i) \longmapsto (l_i, l_i', l_i'')$$

We prove that it is injective: it suffices to consider the case where the values l_i are fixed and θ_i vary. Note that γ_i' and γ_i'' intersect γ_j if and only if $i = j$: hence l_i' and l_i'' depend only on θ_i and not on the other torsion parameters θ_j. Proposition 7.2.10 says that $l_i' = f(\theta_i)$ is strictly convex and Proposition 7.2.4 gives $l_i'' = f(\theta_i + 2\pi)$. A strictly convex proper function $f \colon \mathbb{R} \to \mathbb{R}$ is at most 2 to 1, hence the function

$$\mathbb{R} \longrightarrow \mathbb{R} \times \mathbb{R}$$
$$\theta_i \longmapsto \big(f(\theta_i), f(\theta_i + 2\pi)\big)$$

is injective. Therefore L is injective. $\square$

7.3.3. Collar lemma. The thick-thin decomposition theorem implies that the closed geodesics of length smaller than a Margulis constant ε_2 on a complete hyperbolic surface are simple and have disjoint R-neighbourhoods (see Corollary 4.2.16). When the curves are very short, one may choose R to be very large: this fact is called the *collar lemma* and we prove it directly using elementary tools.

For any number $l > 0$, draw the quadrilateral as in Figure 7.13-(left) and define $f(l)$ to be the distance between its opposite sides l and r.

Exercise 7.3.5. The function $f \colon \mathbb{R}_{>0} \to \mathbb{R}_{>0}$ is strictly decreasing and a homeomorphism. In particular we get $\lim_{l \to 0} f(l) = \infty$. Explicitly, we have

$$\sinh f(l) = \frac{1}{\sinh \frac{l}{2}}.$$

Hint. Put l in vertical position in H^2 and use Lemma 5.2.7. $\square$

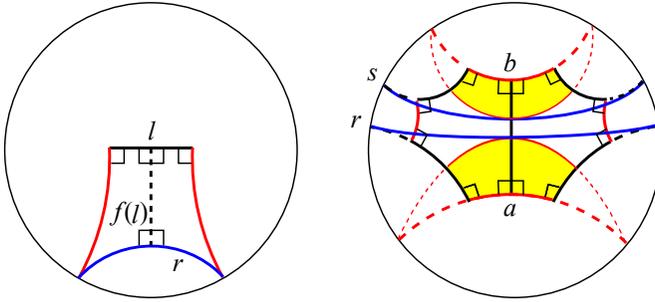The function f is simple to define, and is particularly useful.

Figure 7.13. Pick a segment of length *l* and draw two perpendiculars at the endpoints: this determines a quadrilateral with two ideal vertices; let $f(l)$ be the distance between the opposite sides *l* and *r* (left). A geodesic pair-of-pants is the union of two isometric hexagons, which form in $\mathbb{H}^2$ a right-angled octagon as drawn. The picture shows that the closed boundary geodesics *a* and *b* have disjoint $f(a)$ and $f(b)$-neighbourhoods, coloured here in yellow (right).

Proposition 7.3.6. *Let P be a geodesic pair-of-pants with boundary lengths a, b, and c. The $f(a)$, $f(b)$, and $f(c)$-neighbourhoods of the boundary components form three disjoint collars.*

Proof. Consider two boundary components *a* and *b*. The geodesic pair-of-pants *P* is divided into two isometric hexagons, and we lift them to $\mathbb{H}^2$ where they form a right-angled octagon as in Figure 7.13-(right). The picture shows that the $f(a)$ and $f(b)$-neighbourhoods of *a* and *b* are disjoint. □

Here is the collar lemma.

Lemma 7.3.7 (Collar lemma). *Let $g \geqslant 2$ and S_g have a hyperbolic metric. Disjoint simple closed geodesics $\gamma_1, \ldots, \gamma_k$ of length $l_1, \ldots, l_k$ have disjoint tubular $f(l_i)$-neighbourhoods.*

Proof. We may suppose (by adding more simple closed geodesics if necessary) that the closed geodesics form a pants decomposition, and it suffices to consider two curves that cobound the same pair-of-pants. Proposition 7.3.6 applies. □

We recall from Corollary 4.2.16 that every closed geodesic shorter than ε_2 is simple. Therefore very short closed geodesics are simple and have large disjoint tubular neighbourhoods: the more we shrink the curves, the larger are their neighbourhoods, and hence the larger is the diameter of the surface (the diameter of a metric space is the supremum of the distance of its points). In particular, if on a sequence of closed hyperbolic

surfaces the injectivity radius tends to zero, their diameters must tend to infinity.

Among the many consequences of the collar lemma, we focus on a simple inequality which relates the geometric intersection of simple closed geodesics to their lengths. We denote by $L(\gamma)$ the length of γ.

Corollary 7.3.8. *Let $g \geqslant 2$ and S_g have a hyperbolic metric. Let γ and η be two simple closed geodesics in S_g. The following inequality holds:*

$$L(\eta) \geqslant 2i(\eta, \gamma) \cdot f(L(\gamma)).$$

Proof. The geodesic γ has a tubular $f(L(\gamma))$-neighbourhood. The geodesic η intersects γ in $i(\eta, \gamma)$ points and hence crosses the tubular neighbourhood at least $i(\eta, \gamma)$ times, each with a segment of length $\geqslant 2f(L(\gamma))$. $\square$

7.3.4. A topology for the Teichmüller space. There are various equivalent ways to assign a topology to the Teichmüller space. On the torus T, we have seen that $\mathrm{Teich}(T)$ can be identified with the hyperbolic plane $\mathbb{H}^2$, and the mapping class group acts as isometries on it: we could not hope for a better picture of $\mathrm{Teich}(T)$ and we are fully satisfied.

When $g \geqslant 2$ we could similarly use the Fenchel-Nielsen coordinates and give $\mathrm{Teich}(S_g)$ the topology of $\mathbb{R}^{6g-6}$, but then to be honest we should also check that the topology does not depend on the frame... we prefer to equip the Teichmüller space with an intrinsic topology and then prove that the Fenchel-Nielsen coordinates are homeomorphisms.

We indicate by $\mathscr{S} = \mathscr{S}(S_g)$ the set of all the non-trivial simple closed curves in S_g, considered up to isotopy and orientation reversal (we say that the curves are *unoriented*). Each element $\gamma \in \mathscr{S}$ induces a length function

$$\ell^\gamma \colon \mathrm{Teich}(S_g) \longrightarrow \mathbb{R}_{>0}.$$

We indicate as usual with $\mathbb{R}^{\mathscr{S}}$ the set of all functions $\mathscr{S} \to \mathbb{R}$ and give it the usual product topology (the weakest one such that all the projections are continuous). The natural map

$$\mathrm{Teich}(S_g) \longrightarrow \mathbb{R}^{\mathscr{S}}$$
$$m \longmapsto \big(\gamma \longmapsto \ell^\gamma(m)\big)$$

is injective by Propositions 7.3.3 and 7.3.4. We may hence consider $\mathrm{Teich}(S_g)$ as a subspace of $\mathbb{R}^{\mathscr{S}}$ and assign it the subspace topology. This topology on $\mathrm{Teich}(S_g)$ is the weakest one where the length functions ℓ^γ are continuous. Recall that a topological space is *second-countable* if it has a countable base.

Proposition 7.3.9. *The space $\mathbb{R}^{\mathscr{S}}$ is Hausdorff and second-countable.*

Proof. Every product of Hausdorff spaces is Hausdorff, and every countable product of second-countable spaces is second-countable. $\square$

We recall the following topological fact.

Proposition 7.3.10. *Let $f \colon X \to Y$ be a continuous and proper map between topological spaces. If Y is Hausdorff and second-countable then f is closed.*

Proper maps onto reasonable spaces are closed. If they are also injective, we can obtain more.

Corollary 7.3.11. *Let $f \colon X \to Y$ be a continuous, proper, and injective map between topological spaces. If Y is Hausdorff and second-countable then f is a homeomorphism onto its image.*

We will use this corollary in a moment. Recall that every isometry $\varphi \in \mathrm{Isom}^+(H^2) = \mathbb{PSL}_2(\mathbb{R})$ has a trace $\mathrm{tr}\varphi$ defined only up to sign, whose modulus is > 2 precisely when φ is hyperbolic, see Proposition 2.3.9.

Proposition 7.3.12. *Let $S = \mathbb{H}^2/_\Gamma$ be an orientable hyperbolic surface. Every hyperbolic transformation $\varphi \in \Gamma$ produces a closed geodesic γ in S with*

$$|\mathrm{tr}\varphi| = 2\cosh \frac{L(\gamma)}{2}.$$

Proof. Up to conjugacy we have $\varphi(z) = e^{L(\gamma)} z$. The matrix is

$$\varphi = \begin{pmatrix} e^{\frac{L(\gamma)}{2}} & 0 \\ 0 & e^{-\frac{L(\gamma)}{2}} \end{pmatrix}$$

hence $|\mathrm{tr}\varphi| = 2\cosh \frac{L(\gamma)}{2}$. $\qquad\qquad\square$

In particular, the length of γ depends continuously on the transformation φ. We will use this fact to prove the following. We suppose $g \geqslant 2$.

Proposition 7.3.13. *The Fenchel-Nielsen map*

$$\mathrm{FN} \colon \mathrm{Teich}(S_g) \longrightarrow \mathbb{R}_{>0}^{3g-3} \times \mathbb{R}^{3g-3}$$

is a homeomorphism.

Proof. We consider $\mathrm{Teich}(S_g)$ inside $\mathbb{R}^{\mathscr{S}}$ and examine the inverse

$$\mathrm{FN}^{-1} \colon \mathbb{R}_{>0}^{3g-3} \times \mathbb{R}^{3g-3} \longrightarrow \mathbb{R}^{\mathscr{S}}.$$

We prove that FN^{-1} is continuous. The map FN^{-1} assigns to the parameters (l_i, θ_i) a metric on S_g constructed by attaching right-angled hexagons. Both the hexagons and the attaching maps depend continuously on the parameters (l_i, θ_i) and lift to a tessellation of $\mathbb{H}^2$ into hexagons. Since the decomposition into hexagons varies continuously, its deck transformations vary continuously in $\mathbb{PSL}_2(\mathbb{R})$ and hence the length functions too by Proposition 7.3.12. Therefore FN^{-1} is continuous.

We prove that FN^{-1} is proper. Take a diverging sequence of parameters (l_i, θ_i) (that is, without converging subsequences) in $\mathbb{R}_{>0}^{3g-3} \times \mathbb{R}^{3g-3}$:

we need to show that its image is also a diverging subsequence. The thesis is equivalent to show that the length function of some curve goes to infinity. If $l_i \to +\infty$ for some i we are done. If $l_i \to 0$, the length of any curve intersecting essentially the shrinking curve γ_i goes to infinity by Corollary 7.3.8. It remains to consider the case where the length parameters l_i converge to some non-zero value, but some twist parameter θ_j goes to infinity: in that case the length of any curve intersecting the twisted curve γ_j goes to infinity by Proposition 7.2.10.

Finally, the map FN^{-1} is a homeomorphism onto its image by Corollary 7.3.11. The proof is complete. $\qquad\square$

During the proof we have also discovered the following.

Proposition 7.3.14. *If a sequence $m_i \in \mathrm{Teich}(S_g)$ diverges, there is a $\gamma \in \mathscr{S}$ such that $\ell^\gamma(m_i) \to \infty$ on a subsequence.*

Recall that the action of a topological group G on a topological space X is *continuous* if the action map $G \times X \to X$ is continuous. This implies that G acts on X by homeomorphisms. We give the mapping class group $\mathrm{MCG}(S_g)$ the discrete topology.

Proposition 7.3.15. *The earthquakes and mapping class group actions on the Teichmüller space are continuous.*

Proof. The mapping class group acts on $\mathscr{S}$ by permutations, hence its action on $\mathbb{R}^{\mathscr{S}}$ is continuous. On Fenchel-Nielsen coordinates the earthquake action sends θ to $\theta + \theta'$ and is hence continuous. $\qquad\square$

The immersion in $\mathbb{R}^{9g-9}$ is also a topological embedding.

Proposition 7.3.16. *The injective representation $\mathrm{Teich}(S_g) \hookrightarrow \mathbb{R}^{9g-9}$ furnished by Proposition 7.3.4 is a homeomorphism onto its image.*

Proof. Using Fenchel-Nielsen coordinates the map is clearly continuous. The proof that it is proper is as in Proposition 7.3.13. $\qquad\square$

7.3.5. Surfaces of finite type. We have considered only closed surfaces S_g for simplicity, but most of the arguments exposed in this chapter extend easily to all surfaces $S_{g,b,p}$ of finite type with negative Euler characteristic.

The Teichmüller space $\mathrm{Teich}(S_{g,b,p})$ is the set of all the complete hyperbolic metrics with geodesic boundary, considered up to isometries that are isotopic to the identity. Fenchel-Nielsen coordinates are defined analogously: the surface decomposes into $-\chi(S_{g,b,p}) = 2g + b + p - 2$ pairs-of-pants, and the interior curves of the decomposition are

$$\frac{1}{2}\big(3(2g + b + p - 2) - b - p\big) = 3g + b + p - 3.$$

The Fenchel-Nielsen coordinates are

$$\big(l_1, \dots, l_{3g+b+p-3}, l_1^\partial, \dots, l_b^\partial, \theta_1, \dots, \theta_{3g+b+p-3}\big)$$

where the l_i and l_j^θ are the length parameters of the $3g + b + p - 3$ interior and b boundary curves, and the θ_i are the torsion angles of the interior curves. We get a bijection

$$\mathrm{FN}\colon \mathrm{Teich}(S_{g,b,p}) \longrightarrow \mathbb{R}_{>0}^{3g+2b+p-3} \times \mathbb{R}^{3g+b+p-3}$$

which is a homeomorphism with respect to the natural topology on Te-ichmüller space as a subset of $\mathbb{R}^{\mathscr{S}}$. Therefore the Teichmüller space is homeomorphic to a ball of dimension $-3\chi(S_{g,b,p}) - p$.

For instance, the Teichmüller space of a pair-of-pants is $\mathbb{R}_{>0}^3$ parametrized by the lengths of the boundary geodesics, while that of a thrice-punctured sphere is a point (there is a unique metric).

An alternative description of the Teichmüller space for punctured surfaces, with ideal triangles playing the role of pairs-of-pants, is described in the next section.

7.4. Shear coordinates

The Teichmüller space of a punctured surface may also be parametrized using ideal triangles instead of pairs-of-pants: this viewpoint is maybe a bit simpler, and generalises successfully to dimension three (via ideal tetrahedra). The coordinates that it produces are called *shear coordinates*.

7.4.1. Ideal triangulations.
In dimension two and three it is customary to relax the definition of *triangulation*, originally restricted to simplicial complexes, see Section 1.7.7. We prefer to define a triangulation in a looser sense, as a finite set of triangles glued together by pairing their edges.

Let $\Delta_1, \ldots, \Delta_{2k}$ be an even number of identical copies of the standard oriented 2-simplex. A *triangulation* $\mathcal{T}$ is a partition of the $6k$ edges of the triangles into $3k$ pairs, and for each pair a simplicial isometry between the two edges. The triangulation is *oriented* if the simplicial isometries are orientation-reversing. If we glue the triangles along the isometries we get a compact surface S: we always suppose that S is connected and $\mathcal{T}$ is oriented, hence $S = S_g$ for some $g \geqslant 0$.

The surface S is triangulated with $2k$ triangles, $3k$ edges, and some p vertices. Vertices, edges, and triangles form a cellularisation of S, but not necessarily a simplicial complex: for instance two or three vertices of the same triangle Δ_i can be identified to a single one along the process, as the following exercise shows.

Exercise 7.4.1. Construct a triangulation of the torus with one vertex, three edges, and two triangles.

Let Σ be the non-compact surface obtained by removing the p vertices of the triangulation $\mathcal{T}$ from S: we say that $\mathcal{T}$ is an *ideal triangulation* for Σ. The surface Σ is a *punctured surface*, i.e. $\Sigma = S_{g,0,p}$ with $p \geqslant 1$.
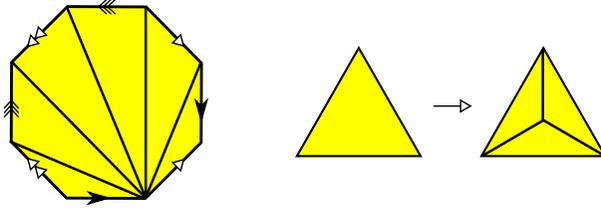
Figure 7.14. The standard representation of S_g by identi-
fying the edges of a $4g$-gon has one vertex; we triangulate
the $4g$-gon and get an ideal triangulation for $S_{g,0,1}$ (left). A
move that modifies a triangulation by adding a new vertex
(right).

Proposition 7.4.2. *We have* $\chi(\Sigma) = -k < 0$. *Every punctured surface
of negative Euler characteristic has an ideal triangulation.*

Proof. We have $\chi(\Sigma) = \chi(S) - p = p - 3k + 2k - p = -k$. An
ideal triangulation for $S_{0,0,3}$ is constructed by attaching Δ_1 and Δ_2 via
the obvious maps. An ideal triangulation for $S_{g,0,1}$ is constructed in Figure
7.14-(left) for all $g \geqslant 1$. An ideal triangulation for $S_{g,0,p}$ with $\chi(S_{g,0,p}) < 0$
and $p \geqslant 2$ is obtained from one of these by increasing p as in Figure 7.14-
(right). $\qquad\square$

7.4.2. Hyperbolic ideal triangulations. Let $\mathcal{T}$ be an oriented trian-
gulation with triangles $\Delta_1, \ldots, \Delta_{2k}$. We substitute every Δ_i with an ideal
hyperbolic triangle (recall that it is unique up to isometry) and pair the
edges with some orientation-reversing isometries. The resulting oriented
punctured surface Σ inherits a hyperbolic structure of area $k\pi$, since an
ideal triangle has area π. We call $\mathcal{T}$ a *hyperbolic ideal triangulation* for Σ.

The edges are infinite lines and the gluing isometry is not unique: in-
deed there is a 1-parameter family of isometries to choose from at every
pair of edges, and the hyperbolic structure depends on that – although the
topology of Σ does not. To encode this dependence, we note that every
edge of an ideal triangle has a preferred *midpoint* defined in Figure 7.15.
The orientation-reversing isometry between two edges of two triangles is
determined by the signed distance d (called *shear*) of their two midpoints
as shown in Figure 7.16-(left). The hyperbolic structure on Σ is deter-
mined by the *shear coordinates* $d = (d_1, \ldots, d_{3k})$ of the $3k$ edges of the
triangulation.

7.4.3. Complete solutions. The hyperbolic structure on Σ may be
incomplete! Recall that a hyperbolic surface is complete if and only if its
universal cover is $\mathbb{H}^2$, so we now look at the universal cover of Σ.
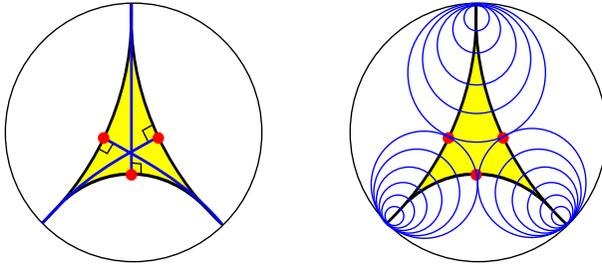
Figure 7.15. On an ideal triangle every edge has a unique perpendicular which ends at the opposite vertex. The base of this perpendicular is the *midpoint* of the edge and the three perpendiculars intersect in a point called *barycenter* (left). There is a unique configuration of three pairwise tangent horocycles centred at the ideal vertices, and their tangency points are the midpoints of the edges (right).
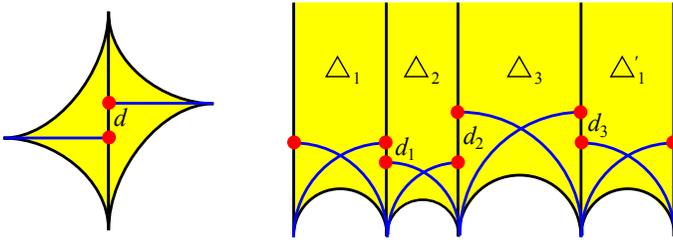

Figure 7.16. The *shear* is the signed distance d between two midpoints after the gluing. By convention the sign is positive $d > 0$ if the points are as shown here, that is an observer that arrives at the midpoint of an edge sees the other midpoint on its left: we need the orientation of $\mathcal{T}$ here (left). We develop a triangulation around a puncture v: here $h = 3$ (right).

Proposition 7.4.3. *The universal cover $\tilde{\Sigma}$ of Σ is isometric to the interior of the intersection of some half-planes in $\mathbb{H}^2$ with disjoint boundary lines (see Figure 7.17). The surface Σ is complete if and only if $\tilde{\Sigma} \cong \mathbb{H}^2$.*

Proof. Recall from Section 3.5.3 the developing map $D \colon \tilde{\Sigma} \to \mathbb{H}^2$. We prove that D is injective.

We can construct D as follows: the hyperbolic ideal triangulation $\mathcal{T}$ of Σ lifts to a hyperbolic ideal triangulation $\tilde{\mathcal{T}}$ of $\tilde{\Sigma}$ with infinitely many triangles. Send a triangle of $\tilde{\mathcal{T}}$ to an arbitrary ideal triangle of $\mathbb{H}^2$ and then develop the map D by attaching subsequently all the triangles of $\tilde{\mathcal{T}}$. At each step the image of D is an ideal polygon and we attach a new triangle
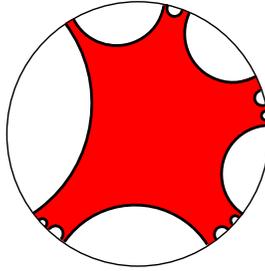
Figure 7.17. The universal covering of Σ is isometric to the interior of a convex set delimited by some lines in $\mathbb{H}^2$, which is the intersection of (possibly infinitely many) half-planes bounded by disjoint lines.

to the exterior of some side of the polygon. Therefore at each step D is injective and hence D is globally injective.

Since D is injective we identify $\tilde{\Sigma}$ with its open image in $\mathbb{H}^2$. Pick a point $x \in \partial\tilde{\Sigma}$. Every neighbourhood of x must intersect infinitely many triangles of $\tilde{\mathcal{T}}$, hence there is a sequence of edges e_i of $\tilde{\mathcal{T}}$ such that $d(e_i, x)$ is monotone decreasing and tends to zero. Since the e_i are disjoint, their endpoints in $\partial\mathbb{H}^2$ are monotone and tend to two points, which determine a line l. We must have $d(l, x) = 0$ and hence $x \in l$ and $l \subset \partial\tilde{\Sigma}$.

We have discovered that $\partial\tilde{\Sigma}$ consists of disjoint lines. This proves that the closure of $\tilde{\Sigma}$ is the intersection of half-planes. A surface Σ is complete if and only if $\tilde{\Sigma}$ is complete, and hence isometric to $\mathbb{H}^2$. $\quad\square$

The hyperbolic structure on Σ depends on the shearing coordinates $d \in \mathbb{R}^{3k}$, and we now want to understand which values of d produce a complete metric. An ideal vertex v of the triangulation $\mathcal{T}$ is adjacent to some h triangles, which we denote by $\Delta_1, \ldots, \Delta_h$ for simplicity although there might be repetitions, and to h edges with some shearing coordinates that we also indicate by $d_1, \ldots, d_h$ for simplicity. The vertex v is a puncture of Σ and let $N(v)$ be a small topological punctured closed disc around v.

Proposition 7.4.4. *The punctured disc $N(v)$ is complete if and only if* $d_1 + \ldots + d_h = 0$.

Proof. We construct a portion of the developing map $D\colon \tilde{\Sigma} \to \mathbb{H}^2$ as follows: we use the half-plane model and send Δ_1 to $\mathbb{H}^2$ with $v = \infty$ as in Figure 7.16-(right), then we develop the triangulation horizontally. In the picture $h = 3$ so we reproduce $\Delta_1, \Delta_2, \Delta_3$ and then another copy Δ_1' of Δ_1. Let $\varphi \in \mathbb{PSL}_2(\mathbb{R})$ be the holonomy isometry that sends Δ_1 to Δ_1'.

If $d_1 + \ldots + d_h = 0$ the midpoints of Δ_1 and Δ_1' are at the same height as in Figure 7.16-(right) and therefore φ is a parabolic map $\varphi\colon z \mapsto z + b$
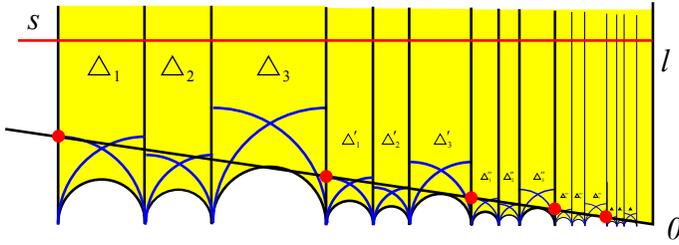
Figure 7.18. When $d_1 + \ldots + d_h \neq 0$, the isometry φ send-
ing Δ_1 to Δ'_1 is hyperbolic, with some axis l which (up to
translations) has endpoints 0 and ∞. Hence $\varphi(z) = e^d z$
and the triangulation develops towards l as shown here.

for some $b \in \mathbb{R}$. Therefore $N(v)$ is a truncation of the cusp $\mathbb{H}^2/_{\langle \varphi \rangle}$, which
is complete.

If $d_1 + \ldots + d_h \neq 0$ the midpoints of Δ_1 and Δ'_1 are at different
heights and therefore φ is not parabolic. Since $\infty \in \mathrm{Fix}(\varphi)$, the isometry
φ is a hyperbolic transformation having some other fixed point in $\mathbb{R}$. Up
to translating everything we may suppose that $\mathrm{Fix}(\varphi) = \{0, \infty\}$, hence
the axis l of φ is the vertical coordinate axis, and $\varphi(z) = e^d z$ with $d = d_1 + \ldots + d_h$. The map D develops towards l as shown in Figure 7.18: we
get $l \subset \partial\tilde{\Sigma}$, so $\tilde{\Sigma} \neq \mathbb{H}^2$ is not complete. $\qquad \square$

Summing up, we have a *completeness equation* of type $d_{i_1} + \ldots + d_{i_h} = 0$ for each of the p ideal vertices of the triangulation.

Corollary 7.4.5. *The hyperbolic structure on Σ is complete if and only
if d satisfies the p completeness equations.*

Proof. Let $N(v)$ be a small punctured closed disc at v for every ideal
vertex v. The closure of $\Sigma \setminus \cup_v N(v)$ is compact. Therefore Σ is complete
if and only if each $N(v)$ is. $\qquad \square$

7.4.4. Shear coordinates for Teichmüller space. The solution space
of the completeness equations is some linear subspace $V \subset \mathbb{R}^{3k}$ of di-
mension at least $3k - p = -3\chi(\Sigma) - p$. Note that this is precisely the
dimension of $\mathrm{Teich}(\Sigma)$, see Section 7.3.5. We have constructed a map

$$\mathrm{sh} \colon V \longrightarrow \mathrm{Teich}(\Sigma).$$

We will prove below that sh is a homeomorphism and hence $\dim V = 3k - p$.
We will use the following.

Proposition 7.4.6. *Let Σ be a complete hyperbolic punctured sur-
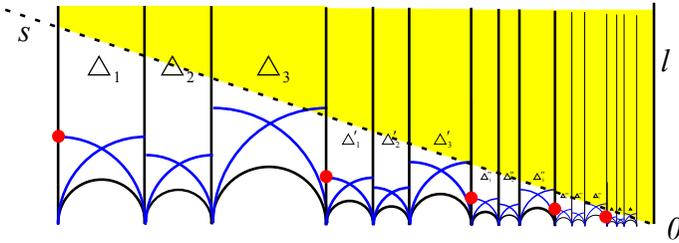face. Every ideal triangulation of Σ is isotopic to a unique hyperbolic ideal
triangulation.*

Figure 7.19. Let s be a Euclidean line at hyperbolic distance R from l. If R is sufficiently small, the line s does not intersect the lower edges of $\Delta_1, \Delta_2, \Delta_3$. Let S be the region lying above s.

Proof. We describe a procedure that *straightens* every ideal triangulation $\mathcal{T}$ of $\Sigma = \mathbb{H}^2/_\Gamma$ to a hyperbolic one.

Truncate the cusps of Σ to get a compact sub-surface. The ideal triangulation $\mathcal{T}$ is now truncated to a decomposition into hexagons, each with three boundary and three interior edges. We lift the truncated cusps and the hexagons to $\mathbb{H}^2$. Truncated cusps lift to horoballs, bounded by horocycles, and hexagons lift to hexagons (the hexagons are not geodesic, only smooth).

An interior edge of a hexagon in $\mathbb{H}^2$ cannot have both its endpoints in the same horocycle, since it would form a bigon with it, and a bigon does not decompose into hexagons (by a Euler characteristic argument). Therefore every interior edge connects two distinct horocycles, centred at two distinct points of $\partial\mathbb{H}^2$. Replace this edge with the geodesic line connecting these two ideal points.

If we do this at every interior edge we get a Γ-invariant hyperbolic ideal triangulation of $\mathbb{H}^2$, which projects to a hyperbolic ideal triangulation for Σ isotopic to $\mathcal{T}$.   □

We can now parametrize $\mathrm{Teich}(\Sigma)$.

Corollary 7.4.7. *The map* sh *is a bijection.*

Proof. Proposition 7.4.6 shows that in every metric the initial triangulation $\mathcal{T}$ straightens to an ideal hyperbolic one and is hence realised by some d, so sh is surjective. Moreover the straightened triangulation is unique and the parameters d_i are obtained intrinsically from it, so sh is also injective.   □

7.4.5. Incomplete metrics. The shear coordinates d may define an incomplete metric on Σ, and we now determine its metric completion $\overline{\Sigma}$.

Recall that every ideal vertex v has a *total shear* $d_v = d_{i_1} + \ldots + d_{i_h}$ which is zero precisely when Σ is complete near v.
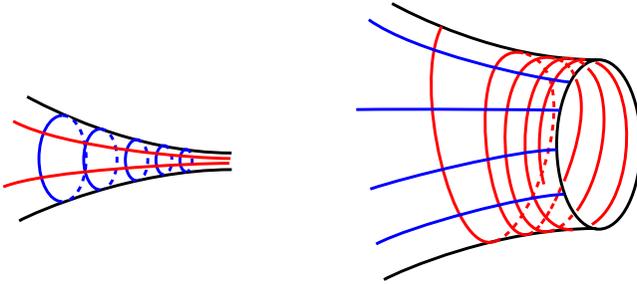
Figure 7.20. Near an ideal vertex v, horocycles (in blue) and edges of the triangulation (in red) are mutually orthogonal. They behave much differently in presence of a cusp (left) or of a geodesic boundary (right): in a cusp, horocycles form parallel circles and edges point toward the cusp; near a geodesic boundary, the horocycle point orthogonally toward the boundary and the edges spin and wind around it indefinitely.

Proposition 7.4.8. *The metric completion $\overline{\Sigma}$ is a hyperbolic surface with geodesic boundary, whose boundary lengths are the absolute values of the total shears of the vertices.*

Proof. Let $N(v)$ be a small closed punctured disc around v. If $d_v = 0$ we already know that $N(v)$ is a truncated cusp. If $d_v \neq 0$ we develop $N(v)$ in $\mathbb{H}^2$ as in Figure 7.19. The hyperbolic transformation is $\varphi(z) = e^{d_v}z$ and the points at distance R from l form a Euclidean line s starting from 0. Pick R small so that s does not intersect the lower edges of $\Delta_1, \ldots, \Delta_h$. Since s is φ-invariant, it is contained in the union of all the developed triangles of Figure 7.19 and hence in $\widetilde{\Sigma}$.

The line s projects to a (non-geodesic) circle in Σ around v, and we can suppose that $\partial N(v)$ is that circle. We have $N(v) = S/_\gamma$ where $S \subset H^2$ is the region lying above s. The completion $\overline{S}$ equals $S \cup l$ and therefore $\overline{N(v)} = \overline{S}/_\gamma$ is an annulus with two boundary components: the original non-geodesic $\partial N(v)$ and a new geodesic $l/_\gamma$, a curve of length $|d_v|$. $\square$

If $d_v \neq 0$, the edges of the triangulation pointing towards v spin around the geodesic boundary as shown in Figure 7.20-(right). The spinning direction induces an orientation on each component of $\partial\overline{\Sigma}$, which depends on the sign of d_v.

Summing up, every shear coordinate $d = (d_1, \ldots, d_{3k}) \in \mathbb{R}^{3k}$ determines a hyperbolic surface $\overline{\Sigma}$ with some p' cusps and b oriented geodesic boundary, where $b + p' = p$ is the number of vertices of the ideal triangulation. The surface $\overline{\Sigma}$ is diffeomorphic to $S_{g,b,p'}$. Let $\text{Teich}^{\text{or}}(S_{g,b,p'})$

denote the Teichmüller space of the surface $S_{g,b,p'}$ equipped with an orientation of the b boundary components: of course this is just 2^b copies of Teich($S_{g,b,p'}$).

Proposition 7.4.9. *The shear coordinates induce a bijection*

$$\mathbb{R}^{3k} \longleftrightarrow \bigsqcup_{b+p'=p} \text{Teich}^{or}(S_{g,b,p'}).$$

Proof. The map sends $d \in \mathbb{R}^{3k}$ to the complete hyperbolic $\overline{\Sigma}$, and we now exhibit an inverse. Let $\overline{\Sigma}$ be a surface of finite type with a hyperbolic structure with p' cusps and b oriented geodesic boundary components, and $\Sigma = \text{int}(\overline{\Sigma})$. We may straighten any ideal triangulation $\mathcal{T}$ of Σ similarly as we did in Proposition 7.4.6, with the only difference that horocycles are replaced by *oriented* geodesic lines as lifts of oriented geodesic boundary components; in the construction, we use the final endpoint of the oriented line instead of the ideal point of the horocycle. We get an ideal triangulation of Σ whose completion is $\overline{\Sigma}$.   □

For instance, $\mathbb{R}^3$ parametrizes altogether all the hyperbolic metrics on the pair of pants, where each boundary component becomes either an oriented geodesic or "degenerates" to a cusp.

7.4.6. References. Most of the material presented here is standard and can be found in many books, starting from Thurston's notes [**56**]. The main reference is Farb – Margalit [**18**], which also contains the proof of Proposition 7.2.10, that appeared originally in a paper of Bestvina, Bromberg, Fujiwara, and Souto [**5**]. The proof of the Collar Lemma is taken from Hubbard [**30**].

Surface diffeomorphisms

We describe in this chapter an analogy between the hyperbolic space $\mathbb{H}^n$ and the Teichmüller space $\mathrm{Teich}(S_g)$ of a closed orientable surface S_g of genus $g \geqslant 1$. The theory, originated from Thurston in the late 1970s, provides a beautiful and powerful framework for the analysis of the geometric and dynamical properties of the surfaces S_g and of their mapping class group $\mathrm{MCG}(S_g)$.

Here is a quick sketch of this analogy. We have already seen that $\mathbb{H}^n$ compactifies to a closed disc $\overline{\mathbb{H}^n}$, that $\mathrm{Isom}(\mathbb{H}^n)$ acts on it, that by Brouwer's theorem every non-trivial isometry φ has a fixed point in $\overline{\mathbb{H}^n}$, and we have called φ *elliptic*, *parabolic*, or *hyperbolic* according to the position of its fixed points.

We construct in this chapter a similar compactification of the open ball $\mathrm{Teich}(S_g)$ to a closed disc. The action of the mapping class group $\mathrm{MCG}(S_g)$ extends to this closed disc, and by Brouwer's theorem every non-trivial element φ of $\mathrm{MCG}(S_g)$ has a fixed point there. According to the position of the fixed points of φ, we say that φ is *finite order*, *reducible*, or *pseudo-Anosov*.

If φ is an isometry for some hyperbolic structure of S_g, it belongs to the first type. Dehn twists belong to the second. The pseudo-Anosov maps are both the most mysterious and the most important, and for their study we need to introduce a wealth of beautiful new technology: *geodesic currents*, *laminations*, and *train tracks*.

8.1. Thurston's compactification

Let S_g be a surface of genus $g \geqslant 2$. Recall that $\mathrm{Teich}(S_g)$ is the Teichmüller space of S_g and $\mathscr{S} = \mathscr{S}(S_g)$ is the set of all non-trivial simple closed curves in S_g, considered up to isotopy and orientation reversal: all the closed curves considered in this chapter are *unoriented*.

Although $\mathrm{Teich}(S_g)$ and $\mathscr{S}$ are very different in nature, we want to compactify both spaces by embedding them in a single bigger space. The model that we have in mind, and that we would like to extend to higher genus surfaces, is the flat torus picture that was painted in Section 7.2.4.

We briefly summarise it. The Teichmüller space $\mathrm{Teich}(T)$ of the torus T is $\mathbb{H}^2$ and it compactifies to $\overline{\mathbb{H}^2}$; the mapping class group $\mathrm{MCG}(T)$ acts

on $\overline{\mathbb{H}^2}$ as a discrete group of isometries; as such, the non-trivial elements of $\mathrm{MCG}(T)$ are divided into three classes (hyperbolic, parabolic, elliptic); the simple closed curves form a dense countable set of rational points in $\partial\mathbb{H}^2 = \mathbb{R} \cup \{\infty\}$.

We now may wonder whether there is an analogous identification between $\mathrm{Teich}(S_g)$ and $\mathbb{H}^{6g-6}$ that transforms $\mathrm{Teich}(S_g)$ into a discrete subgroup of $\mathrm{Isom}(\mathbb{H}^n)$. This is unfortunately not the case, and the compactification of $\mathrm{Teich}(S_g)$ must be constructed from scratch via different methods: we compactify $\mathrm{Teich}(S_g)$ by embedding it in a bigger infinite-dimensional space.

8.1.1. Projective immersion. In Chapter 7 we have used the length functions to construct an embedding

$$i\colon \mathrm{Teich}(S_g) \hookrightarrow \mathbb{R}^{\mathscr{S}}.$$

We know that $\mathrm{Teich}(S_g)$ is homeomorphic to an open ball of dimension $6g-6$, and we want to compactify it in a geometrically meaningful way. A first tentative could be to take its closure in $\mathbb{R}^{\mathscr{S}}$, but this does not work:

Proposition 8.1.1. *The subspace $i(\mathrm{Teich}(S_g))$ is closed in $\mathbb{R}^{\mathscr{S}}$.*

Proof. Inclusion is proper, hence closed by Proposition 7.3.10.   □

We are apparently stuck, so we turn back to our model, hyperbolic space, to get some inspiration. We recall that $\mathbb{H}^n$ is properly embedded in the lorentzian $\mathbb{R}^{n+1}$ as a hyperboloid I^n. To compactify $\mathbb{H}^n$, we may consider its image K^n in $\mathbb{RP}^n$ (the Klein model) and take the closure $\overline{K^n}$ there. We try to mimic this construction, by considering the projective space $\mathbb{P}(\mathbb{R}^{\mathscr{S}})$ with the projection

$$\pi\colon \mathbb{R}^{\mathscr{S}} \setminus \{0\} \longrightarrow \mathbb{P}(\mathbb{R}^{\mathscr{S}}).$$

We first need to check that $\mathrm{Teich}(S_g)$ embeds there.

Proposition 8.1.2. *The composition*

$$\pi \circ i\colon \mathrm{Teich}(S_g) \longrightarrow \mathbb{P}(\mathbb{R}^{\mathscr{S}})$$

is injective.

Proof. Suppose by contradiction that there are two distinct points $m, m' \in \mathrm{Teich}(S_g)$ with $\pi(i(m)) = \pi(i(m'))$; this implies that there is a constant $k > 1$ such that $\ell^\gamma(m) = k \cdot \ell^\gamma(m')$ for all $\gamma \in \mathscr{S}$. That sounds very unlikely, and we now prove that it easily leads to a contradiction.

Let $\gamma_1, \gamma_2 \in \mathscr{S}$ be two curves with $i(\gamma_1, \gamma_2) = 1$. We take $x_0 = \gamma_1 \cap \gamma_2$ as a basepoint for $\pi_1(S_g, x_0)$ and note that the elements $\gamma_2 * \gamma_1$ and $\gamma_2 * \gamma_1^{-1}$ are represented by two more non-trivial simple closed curves in S_g. It is easily checked that the formula

$$\mathrm{tr}(A) \cdot \mathrm{tr}(B) = \mathrm{tr}(AB) + \mathrm{tr}(A^{-1}B)$$

holds for any $A, B \in \mathrm{SL}_2(\mathbb{R})$. Proposition 7.3.12 implies that

$$2\cosh\left(\frac{L(\gamma_1)}{2}\right) \cdot \cosh\left(\frac{L(\gamma_2)}{2}\right) = \cosh\left(\frac{L(\gamma_2 * \gamma_1)}{2}\right) + \cosh\left(\frac{L(\gamma_2 * \gamma_1^{-1})}{2}\right)$$

We have obtained a relation between the lengths of $\gamma_1, \gamma_2, \gamma_2 * \gamma_1$, and $\gamma_2 * \gamma_1^{-1}$ that holds for any hyperbolic metric on S_g. It may be rewritten as:

$$\cosh\left(\tfrac{L(\gamma_1)+L(\gamma_2)}{2}\right)+\cosh\left(\tfrac{L(\gamma_1)-L(\gamma_2)}{2}\right) = \cosh\left(\tfrac{L(\gamma_2*\gamma_1)}{2}\right)+\cosh\left(\tfrac{L(\gamma_2*\gamma_1^{-1})}{2}\right).$$

By contradiction every m'-length is k times a m-length: this equation is hence valid after multiplying every argument by k. It is easy to check that

$$\cosh a + \cosh b = \cosh c + \cosh d, \quad \cosh ka + \cosh kb = \cosh kc + \cosh kd$$

if and only if $\{a, b\} = \{c, d\}$. This leads to a contradiction: the number $L(\gamma_1) + L(\gamma_2)$ is strictly bigger than $L(\gamma_2 * \gamma_1)$ or $L(\gamma_2 * \gamma_1^{-1})$, since $\gamma_2 * \gamma_1$ and $\gamma_2 * \gamma_1^{-1}$ have a non-geodesic representative of length $L(\gamma_1) + L(\gamma_2)$. $\square$

We have embedded $\mathrm{Teich}(S_g)$ in $\mathbb{P}(\mathbb{R}^{\mathscr{S}})$, and we now turn to $\mathscr{S}$.

8.1.2. Thurston's compactification. We now embed $\mathscr{S}$ in $\mathbb{P}(\mathbb{R}^{\mathscr{S}})$. A simple closed curve $\gamma \in \mathscr{S}$ defines a functional $i(\gamma) \in \mathbb{R}^{\mathscr{S}}$ by setting:

$$i(\gamma)(\eta) = i(\gamma, \eta).$$

We have constructed a map $i \colon \mathscr{S} \to \mathbb{R}^{\mathscr{S}}$.

Proposition 8.1.3. *The composition*

$$\pi \circ i \colon \mathscr{S} \longrightarrow \mathbb{P}(\mathbb{R}^{\mathscr{S}})$$

is injective.

Proof. Let $\gamma_1, \gamma_2 \in \mathscr{S}$ be distinct. There is always a curve $\eta \in \mathscr{S}$ with $i(\gamma_1, \eta) \neq 0$ and $i(\gamma_2, \eta) = 0$. (If $i(\gamma_1, \gamma_2) > 0$, simply take $\eta = \gamma_2$. Otherwise, it is an easy exercise.) $\square$

We will now tacitly consider both $\mathrm{Teich}(S_g)$ and $\mathscr{S}$ as subsets of $\mathbb{R}^{\mathscr{S}}$.

Proposition 8.1.4. *The subsets* $\mathrm{Teich}(S_g)$ *and* $\mathscr{S}$ *are disjoint in* $\mathbb{P}(\mathbb{R}^{\mathscr{S}})$.

Proof. For each $\gamma \in \mathscr{S}$ we have $i(\gamma, \gamma) = 0$, while every curve has positive length on any hyperbolic metric. $\square$

We can now state Thurston's compactification theorem.

Theorem 8.1.5. *The closure* $\overline{\mathrm{Teich}(S_g)}$ *of* $\mathrm{Teich}(S_g)$ *in* $\mathbb{P}(\mathbb{R}^{\mathscr{S}})$ *is homeomorphic to the closed disc* D^{6g-6}. *Its interior is* $\mathrm{Teich}(S_g)$ *and its boundary sphere contains* $\mathscr{S}$ *as a dense subset.*

In particular, the closure of $\mathscr{S}$ is homeomorphic to a sphere S^{6g-7}. The proof of Theorem 8.1.5 occupies most of this chapter and will be completed in Section 8.3.18. We will introduce in the process various new geometric objects that play an important role in the topology of manifolds in dimension two and three: *geodesic currents*, *laminations*, and *train tracks*.

For the moment we content ourselves with checking that this projective embedding strategy works at least on the much simpler flat torus case.

8.1.3. The torus. The Teichmüller space of the torus T is described in Section 7.2.4 and needs no further comment; nevertheless, we prove here that the projective embedding strategy works for T, just as a sanity check before approaching the more complicated higher genus surfaces.

Proposition 8.1.6. *The space* $\mathrm{Teich}(T)$ *embeds in* $\mathbb{P}(\mathbb{R}^{\mathscr{S}})$ *and its closure there is homeomorphic to a closed disc* D^2. *The interior of this disc is* $\mathrm{Teich}(T)$ *and its boundary contains* $\mathscr{S}$ *as a dense subset.*

Proof. Everything can be written explicitly by identifying $\mathrm{Teich}(T)$ with the half-plane H^2 and $\mathscr{S}$ with the set $\mathbb{Q} \cup \{\infty\} \subset \partial H^2$. Exercise 6.3.14 and Proposition 7.2.5 give

$$i\left(\frac{p}{q}, \frac{r}{s}\right) = \left|\det\begin{pmatrix} p & r \\ q & s \end{pmatrix}\right| = |ps - qr| = |s| \cdot \left|p - q\frac{r}{s}\right|,$$

$$\ell^{\frac{p}{q}}(z) = \frac{|p + qz|}{\sqrt{\Im z}}.$$

Therefore the images of $\frac{r}{s} \in \mathscr{S}$ and $z \in H^2 = \mathrm{Teich}(T)$ in $\mathbb{P}(\mathbb{R}^{\mathscr{S}})$ are respectively the functionals

$$\frac{p}{q} \longmapsto \left|p - q\frac{r}{s}\right| \text{ if } s \neq 0 \text{ and } |q| \text{ if } s = 0,$$

$$\frac{p}{q} \longmapsto |p + qz|.$$

We could remove the constants $|s|$ and $\sqrt{\Im z}$ because we are considering functionals in $\mathbb{P}(\mathbb{R}^{\mathscr{S}})$ rather than in $\mathbb{R}^{\mathscr{S}}$. We define for all $z \in \overline{H^2}$ the functional

$$f_z : \frac{p}{q} \longmapsto |p + qz| \quad \text{if } z \neq \infty,$$

$$f_\infty : \frac{p}{q} \longmapsto |q|.$$

and we get a continuous immersion $z \mapsto f_z$ of $\overline{H^2}$ into $\mathbb{P}(\mathbb{R}^{\mathscr{S}})$. The immersion is closed because it sends a compact space to a Hausdorff space, hence it is a homeomorphism onto its image. The image is the closure of $\mathrm{Teich}(T)$ and its boundary contains $\mathscr{S}$ as a dense set. $\square$

8.2. Geodesic currents

We will prove Thurston's compactification Theorem 8.1.5 using a slightly different perspective. We are guided by the analogy with the hyperbolic space: when we embed (actually, define) the space $\mathbb{H}^n$ in $\mathbb{R}^{n+1}$, we make an essential use of the lorentzian form $\langle,\rangle$ in $\mathbb{R}^{n+1}$, so that $\mathbb{H}^n$ consists of some points v with $\langle v, v \rangle = -1$, and $\partial\mathbb{H}^n$ may be identified with the light cone rays, that is the rays spanned by vectors v with $\langle v, v \rangle = 0$.

It would be nice to have a similar nice bilinear form $\langle,\rangle$ on $\mathbb{R}^{\mathscr{S}}$, with the property that $\mathsf{Teich}(S_g)$ and $\mathscr{S}$ consist of points with $\langle v, v \rangle = k$ and $\langle v, v \rangle = 0$ respectively, for some fixed $k \neq 0$. Unfortunately we are not able to define such a bilinear form in $\mathbb{R}^{\mathscr{S}}$, and in order to get one we now substitute $\mathbb{R}^{\mathscr{S}}$ with a similar (but more structured) infinite-dimensional space, the space of *geodesic currents*.

In geometric measure theory, a *current* is a measure on some space, which generalises the notion of m-dimensional submanifold in a n-manifold. A *geodesic current* is a similar tool introduced by Bonahon in 1988 specifically designed for geodesics on hyperbolic surfaces.

We will see that the geodesic currents are indeed equipped with a bilinear form, which generalises beautifully both the intersection number $i(\alpha, \beta)$ of curves and the length functions ℓ^γ on the Teichmüller space.

8.2.1. Geodesics. From now on, and through all the rest of this chapter, we will consider geodesics only as subsets, neglecting their parametrisation.

More precisely, let M a complete hyperbolic manifold. We indicate by $\mathscr{G}(M)$ the set of the supports of all the complete non-trivial geodesics $\mathbb{R} \to M$. With a little language abuse, we call an element of $\mathscr{G}(M)$ a *geodesic*.

A geodesic in $\mathscr{G}(M)$ is *closed* if it is the support of a closed geodesic $S^1 \to M$, that is if it is compact; otherwise, it is *open*. We say that a geodesic is *simple* if it has a simple geodesic parametrisation, either as an open geodesic $\mathbb{R} \to M$ or as a closed one $S^1 \to M$.

We are particularly interested in the set $\mathscr{G} = \mathscr{G}(\mathbb{H}^2)$ of lines in $\mathbb{H}^2$. A line is determined by its extremes, hence there is a natural bijection

$$\mathscr{G} \longleftrightarrow \left(\partial\mathbb{H}^2 \times \partial\mathbb{H}^2 \setminus \Delta \right) /_\sim$$

where $\Delta = \{(a, a) \mid a \in \partial\mathbb{H}^2\}$ is the diagonal and $(a, b) \sim (b, a)$. We assign to $\mathscr{G}$ the topology of $\left(\partial\mathbb{H}^2 \times \partial\mathbb{H}^2 \setminus \Delta \right) /_\sim$.

Exercise 8.2.1. The space $\mathscr{G}$ is homeomorphic to an open Möbius strip. The lines intersecting a compact set $K \subset \mathbb{H}^2$ form a compact subset of $\mathscr{G}$.

The isometries of $\mathbb{H}^2$ act naturally on $\mathscr{G}$ by homeomorphisms.

Proposition 8.2.2. *If $S = \mathbb{H}^2/_\Gamma$ is a complete hyperbolic surface there is a natural bijection*

$$\mathscr{G}(S) \longleftrightarrow \mathscr{G}/_\Gamma.$$

Proof. Every geodesic in S lifts to a Γ-orbit of lines in $\mathbb{H}^2$.   □

It is typically more comfortable to lift objects from S to the universal cover $\mathbb{H}^2$, and to study them there: we will often see a geodesic $\gamma \in \mathscr{G}(S)$ as a Γ-orbit of lines in $\mathbb{H}^2$.

We will since now consider only closed surfaces S_g of genus $g \geqslant 2$. We now prove the crucial fact that $\mathscr{G}(S_g)$ depends only mildly on the chosen hyperbolic metric for S_g. An *isomorphism* between two pairs $(\mathscr{G}_1, \Gamma_1)$ and $(\mathscr{G}_2, \Gamma_2)$ of groups Γ_i acting on topological spaces $\mathscr{G}_i$ is an isomorphism $\psi \colon \Gamma_1 \to \Gamma_2$ together with a ψ-equivariant homeomorphism $\mathscr{G}_1 \to \mathscr{G}_2$.

Proposition 8.2.3. *Let $g \geqslant 2$ and $S_g = \mathbb{H}^2/_\Gamma$ have a hyperbolic metric. The pair $(\mathscr{G}, \Gamma)$ does not depend (up to canonical isomorphisms) on the chosen hyperbolic metric.*

Proof. Let m, m' be two hyperbolic structures on S_g, inducing two different coverings $\pi, \pi' \colon \mathbb{H}^2 \to S_g$. The identity map $S_g \to S_g$ lifts to a map $\mathbb{H}^2 \to \mathbb{H}^2$ that extends to an equivariant homeomorphism $\partial \mathbb{H}^2 \to \partial \mathbb{H}^2$ by Theorem 5.2.1. This induces an equivariant homeomorphism $\mathscr{G} \to \mathscr{G}$.   □

Here and in the next sections, the hyperbolic metric on S_g has only an auxiliary role: we need it to identify S_g with $\mathbb{H}^2/_\Gamma$ and to define and study some geometric objects like $\mathscr{G}(S_g) = \mathscr{G}/_\Gamma$, but most of our discoveries will be independent *a posteriori* of the auxiliary hyperbolic metric.

Let $S_g = \mathbb{H}^2/_\Gamma$ be a hyperbolic surface. We note that Γ does *not* act properly discontinuously on $\mathscr{G}$.

Proposition 8.2.4. *The Γ-orbit of a line $l \in \mathscr{G}$ is discrete if and only if l projects to a closed geodesic in S_g.*

Proof. Let $\pi(l) \subset S_g$ be the projection of l in S_g. Since S_g is compact, the projection $\pi(l)$ is *not* a closed geodesic $\Leftrightarrow$ there is a small disc $D \subset S_g$ intersecting $\pi(l)$ into infinitely many distinct segments $\Leftrightarrow$ there is a small disc $D \subset \mathbb{H}^2$ intersecting infinitely many lines of the Γ-orbit of l $\Leftrightarrow$ the Γ-orbit is not discrete.   □

The lines $l \subset \mathbb{H}^2$ that project to closed geodesics in S_g are precisely the axis of the hyperbolic isometries φ in Γ, and the Γ-orbit of one such axis l consists of the axis of all the isometries in Γ conjugate to φ. Two such axis are either incident or ultraparallel by Corollary 4.2.3, and the closed geodesic in S_g is simple if and only if all the distinct axis in the Γ-orbit are ultraparallel.

Proposition 8.2.4 implies in particular that the Γ-orbit of l is discrete only for countably many lines $l \in \mathscr{G}$.

Recall from Corollary 4.1.14 that every homotopically non-trivial closed curve in S_g is homotopic to a unique closed geodesic.

8.2.2. Geodesic currents. We now introduce a measure-theoretical notion which is at first sight unrelated with everything we have seen up to now; we will later show that, on the contrary, it generalises many of the geometric objects that we have encountered in the last pages. The basic notions of measure theory that we will need are summarised in Section 1.3.

Definition 8.2.5. Let $S_g = \mathbb{H}^2/_\Gamma$ have a hyperbolic metric. A *geodesic current* on S_g is a locally finite Γ-invariant Borel measure on $\mathscr{G} = \mathscr{G}(\mathbb{H}^2)$.

We denote by $\mathscr{C} = \mathscr{C}(S_g)$ the set of all the geodesic currents in S_g. It is a subset of the space $\mathscr{M}(\mathscr{G})$ of all the locally finite Borel measures on $\mathscr{G}$, closed with respect to sums and products with non-negative scalars, and it inherits its topology, see Section 1.3.3.

The currents space $\mathscr{C}$ is independent of the auxiliary hyperbolic metric on S_g up to canonical isomorphisms, since the pair $(\mathscr{G}, \Gamma)$ is.

8.2.3. Closed geodesics. We now introduce a fundamental example of geodesic current.

Example 8.2.6 (Closed geodesics). A closed geodesic γ on $S_g = \mathbb{H}^2/_\Gamma$ lifts by Proposition 8.2.4 to a discrete Γ-orbit of lines in $\mathbb{H}^2$. The Dirac measure on this discrete set is locally finite and Γ-invariant, hence it is a geodesic current.

We can therefore interpret every closed geodesic in S_g as a particular geodesic current with discrete support. In particular we get an embedding

$$\mathscr{S} \hookrightarrow \mathscr{C}$$

of the set $\mathscr{S}$ of all (isotopy classes of) unoriented non-trivial simple closed curves in S_g into $\mathscr{C}$. The embedding is defined by taking the geodesic representative of each curve. Recall that the hyperbolic metric plays only an auxiliary role.

The following proposition implies that − conversely − every current supported on a discrete set is a linear combination of closed geodesics.

Proposition 8.2.7. *If $l \in \mathscr{G}$ is an atomic point for a geodesic current μ, that is if $\mu(\{l\}) > 0$, then l projects to a closed geodesic in S_g.*

Proof. Since l is atomic and μ is Γ-invariant and locally finite, the Γ-orbit of l in $\mathscr{G}$ is discrete. We conclude thanks to Proposition 8.2.4. □

It is an important aspect of the theory that atomic points may be only of a very specific type. We remind that many closed geodesics are not simple.
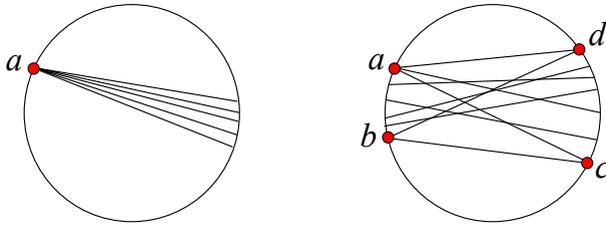
Figure 8.1. A pencil and a box (in the Klein model).

8.2.4. Pencils. We have determined the 0-dimensional subsets of $\mathscr{G}$ that may have positive mass, and we now look at some natural 1-dimensional ones. Let a *pencil* $p \subset \mathscr{G}$ of lines centred at $a \in \partial\mathbb{H}^2$ be some Borel set of lines all having one endpoint at a, as in Figure 8.1-(left).

We note that a pencil may contain at most one axis of some hyperbolic transformation in Γ, because two axis are never asymptotically parallel by Corollary 4.2.3.

Proposition 8.2.8. *Let μ be a geodesic current. The mass of a pencil p is zero, unless it contains the axis of some non-trivial element in Γ.*

Proof. It suffices to consider the case where p contains no axis, and consists of all lines with one endpoint in a and another in some segment $[b, c] \subset \partial\mathbb{H}^2$ not containing a.

As the lines in p point to a, they project in S_g to a narrow beam of lines that go on running forever and exponentially-narrowing in S_g but never close up. There is a disc $D \subset S_g$ intersecting them infinitely many times, hence a lift $D \subset \mathbb{H}^2$ intersects infinitely many disjoint Γ-translates of p. The lines intersecting D form a compact subset of $\mathscr{G}$, hence $\mu(p) = 0$. $\square$

8.2.5. Boxes. We now investigate some natural 2-dimensional subsets of $\mathscr{G}$. Let $a, b, c, d \in \partial\mathbb{H}^2$ be four distinct counterclockwise-ordered points as in Figure 8.1-(right): they determine two disjoint arcs $[a, b]$, $[c, d]$ in $\partial\mathbb{H}^2$ and hence a compact set $B = [a, b] \times [c, d] \subset \mathscr{G}$ consisting of all lines with endpoints in $[a, b]$ and $[c, d]$, see Figure 8.1-(right). We call this compact set B a *box*.

The topological boundary ∂B of a box B has four sides: the four pencils of lines in B with one endpoint in $a, b, c,$ or d. We say that the box B is *generic* if no line $l \in \partial B$ projects to a closed geodesic, *i.e.* no $l \in \partial B$ is the axis of some hyperbolic transformation in Γ. This is indeed a generic condition: if neither of a, b, c, d is the endpoint of some axis, then B is certainly generic, and recall that there are only countably many axis overall.
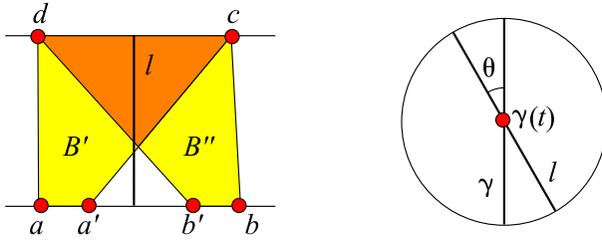
Figure 8.2. The box B containing l and the sub-boxes B', B'' (left). A line $l \in U_\gamma$ parametrised as (t, θ) (right).

We equip the currents space $\mathscr{C}$ with the weak-* topology introduced in Section 1.3.3. We now prove that a converging sequence of currents behave well on generic boxes.

Proposition 8.2.9. *If $\mu_i \rightharpoonup \mu$ is a converging sequence of geodesic currents and B is a generic box, then $\mu_i(B) \to \mu(B)$.*

Proof. By Proposition 1.3.8 it suffices to prove that $\mu(\partial B) = 0$, and this follows from Proposition 8.2.8. □

8.2.6. Rigidity of atoms. We now use boxes to prove that the atomic points in $\mathscr{C}$ are somehow rigid.

Proposition 8.2.10. *The closed geodesics form a discrete subset in $\mathscr{C}$.*

Proof. Let $\mu \in \mathscr{C}$ be a closed geodesic, *i.e* it is the Dirac measure on the Γ-orbit of an axis $l \subset \mathbb{H}^2$ of some hyperbolic transformation $\varphi \in \Gamma$.

Let $B = [a, b] \times [c, d]$ be a small generic box containing l and no other Γ-translate of l, so that $\mu(B) = 1$. Up to substituting φ with φ^{-1} we may suppose that $\varphi([a, b]) = [a', b'] \subset [a, b]$. Consider the sub-boxes $B' = [a, a'] \times [c, d]$ and $B'' = [b', b] \times [c, d]$ shown in Figure 8.2-(left). Since φ contracts $[a, b]$ and expands $[c, d]$, we have

$$(4) \qquad\qquad B \setminus p \subset \bigcup_{i=1}^{\infty} \varphi^i(B' \sqcup B'').$$

where p is the pencil containing l and other lines with one endpoint varying in $[c, d]$. Note that p contains no axis of Γ other than l. If necessary, we move a' and b' slightly farther from a and b to ensure that B' and B'' (and hence all their φ-translates) are generic while (4) still holds.

We have $\mu(B' \sqcup B'') = 0$ and we define an open neighbourhood U of μ in the current space $\mathscr{C}$ as follows:

$$U = \big\{ \eta \in \mathscr{C} \mid \eta(B' \sqcup B'') < \varepsilon, \ \eta(B) > 1 - \varepsilon \big\}$$

for some fixed $0 < \varepsilon < 1$. This is indeed an open set because B, B', and B'' are generic. We now prove that U contains no closed geodesic except μ.

By contradiction, suppose U contains a closed geodesic η. Since $\eta(B) > 1 - \varepsilon > 0$, at least one atom l' of η is contained in B. If $l' = l$ then $\eta = \mu$ and we are done. Otherwise, the line l' is contained in $B \setminus p$ and hence in $\varphi^i(B' \sqcup B'')$ for some i. This is impossible since this set has measure $\varepsilon < 1$.   □

We have classified the elements in $\mathscr{C}$ supported on discrete sets, and it is now due time to introduce some geodesic currents supported on the whole of $\mathscr{G}$.

8.2.7. The Liouville measure. There is a continuous measure on $\mathscr{G} = \mathscr{G}(\mathbb{H}^2)$ that is invariant under the action of the whole isometry group $\text{Isom}(\mathbb{H}^2)$. This measure is called the *Liouville measure*, and is defined as follows.

Let $\gamma \colon \mathbb{R} \to \mathbb{H}^2$ be a geodesic parametrised by arc length, and $U_\gamma \subset \mathscr{G}$ be the open set consisting of all the lines intersecting γ, except γ itself. We can parametrise U_γ via the homeomorphism

$$\mathbb{R} \times (0, \pi) \to U_\gamma$$

that sends (t, θ) to the line l that intersects γ at the point $\gamma(t)$ with angle θ, see Figure 8.2-(right). We define a volume 2-form on U_γ by setting:

$$L_\gamma = \frac{1}{2} \sin\theta \, dt \wedge d\theta.$$

Since $\mathscr{G}$ is non-orientable, we cannot hope to define a global area form on $\mathscr{G}$. However all these local forms match up to sign and hence give rise to a measure.

Proposition 8.2.11. *The charts U_γ form a differentiable atlas for $\mathscr{G}$. The 2-forms L_γ match up to sign and hence define a measure L on $\mathscr{G}$.*

Proof. Every line in $\mathbb{H}^2$ intersects some other line, hence the charts cover $\mathscr{G}$. We consider a line $r \in U_\gamma \cap U_{\gamma'}$. The charts U_γ and $U_{\gamma'}$ have parametrisations (t, θ) and (t', θ') and 2-forms

$$L_\gamma = \frac{1}{2} \sin\theta dt \wedge d\theta, \quad L_{\gamma'} = \frac{1}{2} \sin\theta' dt' \wedge d\theta'.$$

We consider the jacobian $J = \frac{\partial(t', \theta')}{\partial(t, \theta)}$ and recall that

$$dt' \wedge d\theta' = \det J \cdot dt \wedge d\theta$$

so we need to show that

(5) $$\det J = \pm \frac{\sin\theta}{\sin\theta'}.$$

We can reparametrise γ at our please, since this may change L_γ only by a sign. We may suppose that γ and γ' are asymptotically parallel, because
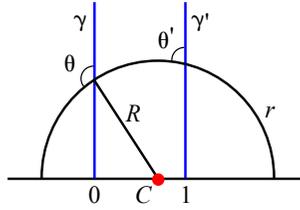
Figure 8.3. The two asymptotic lines γ, γ' in the half-space model H^2 and the line r intersecting both.

any two lines γ, γ' intersecting r are connected by a path $\gamma = \gamma_1, \ldots, \gamma_k = \gamma'$ of lines intersecting r, such that any two subsequent lines γ_i, γ_{i+1} are asymptotically parallel (actually $k = 3$ suffices).

We represent the lines γ, γ', and r in the half-plane model $H^2 \subset \mathbb{R}^2$ as in Figure 8.3. Up to isometries and reparametrisations we have

$$\gamma(t) = (0, e^t), \qquad \gamma'(t') = (1, e^{t'}).$$

The line r is a Euclidean half-circle with center at some point $C \in \mathbb{R}$ and with some radius $R > 0$. It intersects γ and γ' at some heights e^t and $e^{t'}$, with some angles θ and θ'. We now calculate the determinant of $J = \frac{\partial(t',\theta')}{\partial(t,\theta)}$. Figure 8.3 shows that

$$R = \frac{e^t}{\sin\theta} = \frac{e^{t'}}{\sin\theta'}$$

$$C = -e^t \cot\theta, \qquad 1 - C = e^{t'} \cot\theta'.$$

Now

$$\frac{\partial(C, R)}{\partial(t, \theta)} = \begin{pmatrix} -e^t \cos\theta \sin^{-1}\theta & e^t \sin^{-2}\theta \\ e^t \sin^{-1}\theta & -e^t \cos\theta \sin^{-2}\theta \end{pmatrix}$$

whose determinant is $-e^{2t} \sin^{-1}\theta = -R\sin\theta$, and we get the same formula for $\frac{\partial(C,R)}{\partial(t',\theta')}$. Therefore

$$\det \frac{\partial(t', \theta')}{\partial(t, \theta)} = \det \frac{\partial(t', \theta')}{\partial(C, R)} \cdot \det \frac{\partial(C, R)}{\partial(t, \theta)} = (-R\sin\theta')^{-1} \cdot (R\sin\theta) = \frac{\sin\theta}{\sin\theta'}$$

and we finally obtain (5), as required.   □

The measure L on $\mathscr{G}$ is called the *Liouville measure* and is clearly invariant by the action of $\mathrm{Isom}(\mathbb{H}^2)$. The renormalising factor $\frac{1}{2}$ in the definition was chosen to get the following property, which nicely characterises L.

Proposition 8.2.12. *Let $s \subset \mathbb{H}^2$ be a geodesic segment of length l. The lines in $\mathbb{H}^2$ intersecting s form a set of measure l.*

Proof. The set has measure

$$\int_0^\pi \int_0^l \frac{1}{2} \sin\theta \, dt \, d\theta = l \int_0^\pi \frac{1}{2} \sin\theta \, d\theta = l.$$

The proof is complete. □

The Liouville measure L is supported on the whole of $\mathscr{G}$.

Exercise 8.2.13. The Liouville measure of a box $B = [a, b] \times [c, d]$ is

$$L(B) = \big| \log \beta(a, b, c, d) \big|$$

where β is the cross-ratio of the four points.

8.2.8. The Liouville current. Let now $S_g = \mathbb{H}^2/\Gamma$ be equipped with a hyperbolic metric. The Liouville measure on $\mathscr{G}$ is $\mathrm{Isom}(\mathbb{H}^2)$-invariant: in particular it is Γ-invariant and hence defines a current $L \in \mathscr{C}(S_g)$, called the *Liouville current*.

It is important to note that the space of currents $\mathscr{C} = \mathscr{C}(S_g)$ does not depend on the given hyperbolic metric for S_g, but the Liouville current does! Every metric $m \in \mathrm{Teich}(S_g)$ induces a Liouville current $L_m \in \mathscr{C}$, and in this way we get a *Liouville map*

$$\mathrm{Teich}(S_g) \longrightarrow \mathscr{C}$$

that sends m to L_m. We will soon see that this map is injective: as promised, we have mapped both $\mathscr{S}$ and $\mathrm{Teich}(S_g)$ inside the currents space $\mathscr{C}$.

Remark 8.2.14. We constructed the Liouville current L_m by assigning to $\mathscr{G}$ a differentiable atlas – hence a smooth structure – and a 2-form defined up to sign. We note that the smooth structure on $\mathscr{G}$ also depends crucially on m, because the boundary extension of Theorem 5.2.1 is not guaranteed to be a diffeomorphism.

The crucial feature that makes $\mathscr{C}$ preferable to $\mathbb{P}(\mathbb{R}^{\mathscr{S}})$ as a comfortable ambient space for both $\mathscr{S}$ and $\mathrm{Teich}(S_g)$ is the existence of a nice bilinear form in $\mathscr{C}$ that extends both the length and the geometric intersection of closed geodesics. We now introduce this bilinear form.

8.2.9. The projective frame bundle. We denote by $\mathscr{I} \subset \mathscr{G} \times \mathscr{G}$ the open subset consisting of all the pairs of *incident* distinct lines in $\mathbb{H}^2$. We give $\mathscr{I}$ the topology induced by $\mathscr{G} \times \mathscr{G}$, hence $\mathscr{I}$ is an open topological 4-manifold. Since two distinct incident lines intersect in a single point $p \in \mathbb{H}^2$, the set $\mathscr{I}$ can be interpreted as the set of triples (p, l_1, l_2) with $p \in \mathbb{H}^2$ and l_1, l_2 two distinct vector lines in the tangent plane $T_p\mathbb{H}^2$.

Let now $S_g = \mathbb{H}^2/\Gamma$ be equipped with a hyperbolic metric. The action of Γ on $\mathscr{G}$ is not properly discontinuous, but the diagonal action of Γ on $\mathscr{I}$ is, as the following shows.

Proposition 8.2.15. *The map $\mathscr{I} \to \mathscr{I}/\Gamma$ is a topological covering.*

Proof. The group Γ acts freely and properly discontinuously on $\mathbb{H}^2$, and hence it does so *a fortiori* on the triples (p, l_1, l_2). □

In particular $\mathscr{I}/\Gamma$ is a four-manifold, and it is naturally a bundle over S_g, whose fibre above $p \in S_g$ consists of the ordered pairs of distinct vector lines l_1, l_2 in $T_p S_g$. Recall that the *frame bundle* on a manifold M is a bundle whose fibre above $p \in M$ is the set of all *frames* (i.e. basis) in $T_p M$. The space $\mathscr{I}/\Gamma$ can be seen as some projective quotient of the frame bundle on S_g.

8.2.10. Intersection form. Two geodesic currents $\alpha, \beta \in \mathscr{C} = \mathscr{C}(S_g)$ induce a Γ-invariant product measure $\alpha \times \beta$ on $\mathscr{G} \times \mathscr{G}$ and hence on $\mathscr{I}$. This measure descends via the covering $\mathscr{I} \to \mathscr{I}/\Gamma$ to a measure on $\mathscr{I}/\Gamma$ which we still indicate by $\alpha \times \beta$, see Section 1.5.4.

Definition 8.2.16. The *intersection* $i(\alpha, \beta)$ of two geodesic currents is the total volume of $\mathscr{I}/\Gamma$ in the measure $\alpha \times \beta$.

The intersection $i(\alpha, \beta)$ does not depend on the auxiliary hyperbolic metric for S_g. The finiteness of $i(\alpha, \beta)$ is not immediate since $\mathscr{I}/\Gamma$ is not compact, and it deserves a proof.

Proposition 8.2.17. *The intersection $i(\alpha, \beta)$ is finite.*

Proof. Let D be a compact fundamental domain for $S_g = \mathbb{H}^2/\Gamma$. Since D is compact, the lines intersecting D form a compact subset $X \subset \mathscr{G}$ which has finite measures $\alpha(X)$ and $\beta(X)$. The projection sends $(X \times X) \cap \mathscr{I}$ surjectively onto $\mathscr{I}/\Gamma$. Therefore $i(\alpha, \beta) = (\alpha \times \beta)(\mathscr{I}/\Gamma) < \alpha(X) \cdot \beta(X)$. □

The form i is clearly bilinear and symmetric.

8.2.11. Geometric intersection and length of closed geodesics. We now show that this abstract-looking intersection form generalises both the geometric intersection and the length of simple closed geodesics. We consider as usual the set $\mathscr{S}$ of simple closed curves as a subset of $\mathscr{C}$. The following proposition explains why we employed the notation i for the intersection form of two geodesic currents.

Proposition 8.2.18. *If $\alpha, \beta \in \mathscr{S}$, the number $i(\alpha, \beta)$ is the geometric intersection of the simple closed curves α and β.*

Proof. Represent α and β as simple closed geodesics. The measure $\alpha \times \beta$ is the Dirac measure with support the pairs of incident lines in $\mathbb{H}^2$ that cover respectively α and β. The Γ-orbits of these pairs are in natural bijection with the transverse intersection points in $\alpha \cap \beta$. Hence the volume of $\mathscr{I}/\Gamma$ is the cardinality of $\alpha \cap \beta$, except when $\alpha = \beta$ and in this case we get zero. □

We note in particular that $i(\alpha, \alpha) = 0$ for every $\alpha \in \mathscr{S}$, hence the set $\mathscr{S}$ is contained in the "light cone" consisting of all the currents α with $i(\alpha, \alpha) = 0$. (Of course $i(\alpha, \alpha) \geqslant 0$ for every geodesic current α.) We now consider the Liouville current $L_m \in \mathscr{C}$ determined by some hyperbolic metric m on S_g.

Proposition 8.2.19. *If $\alpha \in \mathscr{S}$ then $i(L_m, \alpha) = \ell^\alpha(m)$ is the length of the geodesic representative of α in the metric m.*

Proof. Represent α as a simple closed geodesic in the metric m, of length $\ell^\alpha(m)$. The measure $L_m \times \alpha$ has its support on the pairs (l, l') of incident lines where l is arbitrary and l' is a lift of α.

Any segment $s' \subset l'$ of length $\ell^\alpha(m)$ in a fixed lift l' of α is a fundamental domain for the action of Γ on the lifts of α. Therefore $i(L_m, \alpha)$ is the volume of the pairs (l, l') where l' is fixed, and l is arbitrary and intersects s'. By Proposition 8.2.12 these pairs have volume $\ell^\alpha(m)$. $\square$

The intersection form i on $\mathscr{C}$ generalises both the geometric intersection of curves and the length functions on Teichmüller space, two apparently unrelated objects! As promised, we can easily deduce that $\mathrm{Teich}(S_g)$ embeds in $\mathscr{C}$.

Corollary 8.2.20. *The Liouville map $\mathrm{Teich}(S_g) \to \mathscr{C}$ is injective.*

Proof. If $m \neq m'$ there is a curve $\gamma \in \mathscr{S}$ with $\ell^\gamma(m) \neq \ell^\gamma(m')$ by Proposition 7.3.4 and thus $i(L_m, \gamma) \neq i(L_{m'}, \gamma)$. Hence $L_m \neq L_{m'}$. $\square$

We will since now consider both $\mathrm{Teich}(S_g)$ and $\mathscr{S}$ as subsets of $\mathscr{C}$. We know the geometric meaning of the intersection of two curves, and of a curve and a hyperbolic metric. What is the intersection of two hyperbolic metrics? When they coincide, we get a positive constant.

Proposition 8.2.21. *If $m \in \mathrm{Teich}(S_g)$ we have $i(L_m, L_m) = -\pi^2 \chi(S_g)$.*

Proof. The metric m produces a smooth structure on $\mathscr{G}$ and hence on $\mathscr{I}$, and L_m is induced by a Γ-invariant 2-form on $\mathscr{G}$ defined up to sign, that we also denote as L_m. We need to integrate the 4-form $L_m \times L_m$ on $\mathscr{I}/\Gamma$.

Recall that $\mathscr{I}/\Gamma$ may be interpreted as a bundle over S_g. We make a first-order computation on a very small region of S_g, which we may suppose to be Euclidean: a very thin rectangle R in S_g of sides l and a with $a \ll l$. We compute the volume of the portion of bundle lying above R.

We integrate $L_m \times L_m$ on all the pairs of segments intersecting in some point in R, and we consider only segments with both endpoints on the big l sides, neglecting the small a sides. A segment meeting the l sides at angle θ has length $L(\theta) = a/\sin\theta$. By Proposition 8.2.12 the $L_m \times L_m$ volume of the portion of $\mathscr{I}$ above R is (at a first order) equal to

$$\int_0^\pi \int_0^l \frac{1}{2} \sin\theta \cdot L(\theta) \, dt \, d\theta \sim \frac{1}{2} l \int_0^\pi \sin\theta \cdot \frac{a}{\sin\theta} = \frac{\pi}{2} l a.$$

The first-order contribution of a small region is hence $\frac{\pi}{2}$ times its area. Since the volume is induced by a differentiable form, by taking the limit we find that the contribution of any region is precisely $\frac{\pi}{2}$ its area and we conclude by the Gauss-Bonnet theorem. $\square$

We have already noted that i is symmetric and bilinear. As we mentioned, there is an evident analogy between the embeddings $\mathrm{Teich}(S_g) \hookrightarrow \mathscr{C}$ and $\mathbb{H}^n \hookrightarrow \mathbb{R}^{n+1}$, since there is a bilinear form on both spaces $\mathscr{C}$ and $\mathbb{R}^{n+1}$, and $\mathrm{Teich}(S_g)$ is also contained in a "hyperboloid" consisting of vectors v with $i(v, v) = k$ for some fixed $k \neq 0$. On the other hand $\mathscr{S}$ is contained in the "light cone" formed by all vectors v with $i(v, v) = 0$.

We may compactify $\mathbb{H}^n$ by projecting the hyperboloid to the Klein model $K^n \subset \mathbb{RP}^n$. Now K^n is an open disc, whose boundary is the image of the light cone. We would like to apply the same compactification strategy to $\mathrm{Teich}(S_g)$. Before doing that, we need to prove a slightly technical fact: that the bilinear form i is continuous.

8.2.12. Continuity of the intersection form. Let $S_g = \mathbb{H}^2/\Gamma$ be a closed hyperbolic surface. The following theorem is not obvious because $\mathscr{I}$ is not compact: the proof relies on the fact that closed geodesics form a discrete rigid set in $\mathscr{C}$ and therefore atomic points cannot "escape to infinity" on a converging sequence of currents.

Theorem 8.2.22. *The form* $i \colon \mathscr{C} \times \mathscr{C} \to \mathbb{R}$ *is continuous.*

Proof. Let β_1, β_2 be two currents converging to α_1, α_2. We need to prove that $i(\beta_1, \beta_2)$ converges to $i(\alpha_1, \alpha_2)$. The bundle $\mathscr{I}/\Gamma$ is not compact and we need to control that no mass escapes to infinity: the infinity here is the diagonal $\Delta \subset \mathscr{G} \times \mathscr{G}$, so our aim is to cover Δ with small boxes and prove that they contribute very little to $i(\beta_1, \beta_2)$. For any box B, we define
$$\Psi(B) = \big((B \times B) \cap \mathscr{I}\big)/\Gamma.$$
We show that for every $\varepsilon > 0$ there are finitely many small generic boxes B_i whose Γ-translates cover $\mathscr{G}$, such that
$$(\beta_1 \times \beta_2)(\cup_i \Psi(B_i)) < \varepsilon$$
as soon as β_1, β_2 are sufficiently close to α_1, α_2. Since the Γ-translates of $\cup_i(B_i \times B_i)$ form an open neighbourhood of Δ, the set $(\mathscr{I}/\Gamma) \setminus (\cup_i \Psi(B_i))$ is compact with zero-measure boundary (because the boxes B_i are generic) so its contribution to $i(\beta_1, \beta_2)$ tends to that to $i(\alpha_1, \alpha_2)$ if β_1, β_2 are sufficiently close, and the theorem is proved.

It remains to construct the boxes B_i. We first fix finitely many generic boxes B_i whose Γ-translates cover $\mathscr{G}$. Let $K > 0$ be bigger than the total α_1 and α_2-mass of the boxes.

We now pick a small number $0 < \varepsilon' < \varepsilon/2K$ and subdivide each box B_i into finitely many generic sub-boxes (which we still call B_i) such that the following holds for each i and each $j = 1, 2$:

(1) either $\alpha_j(B_i) < \varepsilon'$, or
(2) there is an atomic point $l \in B_i$ with $\alpha_j(l) > \varepsilon'$ and $\alpha_j(B_i \setminus l) < \varepsilon'$.

If the second case holds for both $j = 1, 2$, we also require $l \in B_i$ to be the same line for $j = 1, 2$. We note that the total α_1 and α_2-mass of the boxes is still smaller than K after the subdivision. We have

$$(\beta_1 \times \beta_2)(\Psi(B)) \leqslant \beta_1(B) \cdot \beta_2(B)$$

so if β_1, β_2 are sufficiently close to α_1, α_2 the boxes B_i that are of the first kind for at least one value of $j = 1, 2$ contribute to $i(\beta_1, \beta_2)$ less than $2\varepsilon' K < \varepsilon$.

We are left to consider the boxes B_i containing a line l whose α_1 and α_2-mass are both bigger than ε'. This is a potentially bad situation since the point $(l, l) \in \mathscr{G} \times \mathscr{G}$ is atomic and contained in the frontier of $\mathscr{I}$, and could enter inside $\mathscr{I}$ abruptly for some arbitrarily small perturbations of α_1 and α_2. We show that this cannot happen because atomic points are rigid, as noted in Proposition 8.2.10.

As in the proof of Proposition 8.2.10, we set $B_i = B = [a, b] \times [c, d]$ and construct $B' = [a, a'] \times [c, d]$, $B'' = [b', b] \times [c, d]$ whose φ-translates cover $B \setminus p$ where p is a pencil containing l. We note that

$$(\beta_1 \times \beta_2)(\Psi(B)) \leqslant \beta_1(B' \sqcup B'') \cdot \beta_2(B) + \beta_2(B' \sqcup B'') \cdot \beta_1(B)$$

is smaller than $\varepsilon'(\beta_1(B) + \beta_2(B))$ and hence by summing on the B_i we get again a contribution smaller than $2\varepsilon' K < \varepsilon$. $\square$

8.2.13. Filling geodesic currents. We now define an interesting class of currents, whose importance will be evident in the next section.

We say that a geodesic current $\alpha \in \mathscr{C}(S_g)$ *fills* the surface $S_g = \mathbb{H}^2/_\Gamma$ if every line in $\mathbb{H}^2$ intersects transversely at least one line in the support of α. For instance, a Liouville measure fills S_g since its support is the whole of $\mathscr{G}$. We say that k closed geodesics $\gamma_1, \ldots, \gamma_k$ *fill* S_g if the geodesic current $\gamma_1 + \ldots + \gamma_k$ does.

Proposition 8.2.23. *Let $\gamma_1, \ldots, \gamma_k$ be closed geodesics. If $S_g \setminus (\gamma_1 \cup \cdots \cup \gamma_k)$ consists of polygons, the curves fill S_g.*

Proof. Every geodesic in S_g intersects these curves. $\square$

As an example, we can pick two multicurves η and μ in S_g that intersect transversely forming only polygons but no bigons. By the bigon criterion η and μ are in minimal position and by Proposition 6.3.24 their geodesic representatives have the same configuration as $\eta \cup \mu$, so they fill S_g.

It is a bit more difficult to construct a single (non-simple) closed geodesic that fills S_g. The following remark is straightforward.

Remark 8.2.24. Let α and β be currents. We have $i(\alpha, \beta) > 0$ if and only if there are two distinct intersecting lines in the supports of α and β.

Corollary 8.2.25. *If α fills S_g then $i(\alpha, \beta) > 0$ for every non-trivial $\beta \in \mathscr{C}$.*

8.2.14. A compactness criterion. We now state a simple and general compactness criterion for subsets of $\mathscr{C}$ that has various nice (and apparently unrelated) geometric consequences.

Proposition 8.2.26 (Compactness criterion). *If $\alpha \in \mathscr{C}$ fills S_g, the set of all $\beta \in \mathscr{C}$ with $i(\alpha, \beta) \leqslant M$ is compact, for all $M > 0$.*

Proof. Let $C \subset \mathscr{C}$ be the set of all β with $i(\alpha, \beta) \leqslant M$. It is closed because i is continuous.

Let l be a line in $\mathbb{H}^2$. By hypothesis there is another line l' in the support of α which intersects l in a point. Let B', B be two sufficiently small boxes neighbourhoods of l', l in $\mathscr{G}$, so that $B' \times B \subset \mathscr{I}$ and $B' \times B$ is mapped injectively into $\mathscr{I}/_\Gamma$. If $\beta \in C$ we have

$$\alpha(B')\beta(B) = (\alpha \times \beta)(B' \times B) \leqslant (\alpha \times \beta)(\mathscr{I}/_\Gamma) = i(\alpha, \beta) \leqslant M.$$

Therefore every line l in $\mathbb{H}^2$ has a box neighbourhood B such that

$$\beta(B) \leqslant K_l \quad \forall \beta \in C$$

for some constant $K_l = M/\alpha(B')$ that depends only on l. The set C is relatively compact by Theorem 1.3.9, and hence compact since it is closed. The proof is complete. □

We now state some corollaries. The first concerns the immersion of Teich(S_g) in $\mathscr{C}$.

Corollary 8.2.27. *The Liouville map* Teich(S_g) $\hookrightarrow \mathscr{C}$ *is proper and a homeomorphism onto its image.*

Proof. We prove that the map is proper: if $m_j \in$ Teich(S_g) is a diverging sequence of metrics, we know from Proposition 7.3.14 that there is a closed curve $\gamma \in \mathscr{S}$ such that $\ell^\gamma(m_j) = i(m_j, \gamma) \to \infty$ on a subsequence. Since i is continuous, the sequence m_j diverges also in $\mathscr{C}$.

We denote the Liouville map by L. We should now prove that L is continuous, but we prefer to consider the inverse map $L^{-1} \colon L(\text{Teich}(S_g)) \to$ Teich(S_g). The map L^{-1} is continuous because i is and Teich(S_g) has the weakest topology where the length functions are continuous. We show that L^{-1} is proper. Let $\gamma_1, \ldots, \gamma_k$ be simple closed curves that fill S_g. If $L(m_i)$ is a diverging sequence, by Proposition 8.2.26 we have $i(L(m_i), \sum_t \gamma_t) \to \infty$ and hence $i(L(m_i), \gamma_t) \to \infty$ for some t. Therefore m_i is divergent also in Teich(S_g).

Now L^{-1} is continuous and proper and hence a homeomorphism by Corollary 7.3.11. □

A second immediate corollary is a general compactness criterion for the Teichmüller space.

Corollary 8.2.28. *Let $\gamma_1, \ldots, \gamma_k$ be some closed geodesics that fill S_g. The metrics $m \in \text{Teich}(S_g)$ with $\ell^{\gamma_i}(m) \leqslant M$ form a compact subset of $\text{Teich}(S_g)$, for all $M > 0$.*

We can similarly deduce the following.

Corollary 8.2.29. *Let $\gamma_1, \ldots, \gamma_k$ be simple closed curves that fill S_g. For every M there are only finitely many $\alpha \in \mathscr{S}$ with $i(\gamma_i, \alpha) < M$ for all i.*

Exercise 8.2.30. Use the compactness criterion to re-prove that on a hyperbolic closed surface there are only finitely many closed geodesics of bounded length.

We are finally ready to construct Thurston's compactification of the Teichmüller space, using Bonahon's geodesic currents.

8.2.15. Projective currents. As usual, we pick a hyperbolic closed surface $S_g = \mathbb{H}^2/_\Gamma$. The currents space $\mathscr{C}$ is equipped with a multiplication by positive scalars, hence we can define its projectivisation

$$\pi \colon \mathscr{C} \setminus 0 \longrightarrow \mathbb{P}\mathscr{C}$$

where $\mathbb{P}\mathscr{C} = (\mathscr{C} \setminus 0)/_\sim$ with $\alpha \sim \lambda\alpha$ for all $\lambda > 0$. We give $\mathbb{P}\mathscr{C}$ the quotient topology.

Proposition 8.2.31. *The space $\mathbb{P}\mathscr{C}$ is compact.*

Proof. Pick an $\alpha \in \mathscr{C}$ that fills S_g. By the compactness criterion the set $C = \{\beta \in \mathscr{C} \mid i(\alpha, \beta) = 1\}$ is compact. By Corollary 8.2.25 we have $i(\alpha, \beta) > 0$ for all β, hence $\lambda\beta \in C$ for some $\lambda > 0$. Therefore $\pi(C) = \mathbb{P}\mathscr{C}$ and $\mathbb{P}\mathscr{C}$ is compact.   □

We now want to embed both $\text{Teich}(S_g)$ and $\mathscr{S}$ in $\mathbb{P}\mathscr{C}$.

Proposition 8.2.32. *The composition $\mathscr{S} \to \mathscr{C} \setminus 0 \to \mathbb{P}\mathscr{C}$ is injective.*

Proof. Let $\gamma_1, \gamma_2 \in \mathscr{S}$ be distinct. There is always a closed curve $\eta \in \mathscr{S}$ with $i(\gamma_1, \eta) \neq 0$ and $i(\gamma_2, \eta) = 0$.   □

Proposition 8.2.33. *The composition $\text{Teich}(S_g) \to \mathscr{C} \setminus 0 \to \mathbb{P}\mathscr{C}$ is injective and a homeomorphism onto its image.*

Proof. We see $\text{Teich}(S_g)$ already properly embedded in $\mathscr{C}$. Since $i(m, m) = -\pi^2 \chi(S_g)$ is constant on $\text{Teich}(S_g)$, the composition is injective. The restriction $\pi \colon \text{Teich}(S_g) \to \pi(\text{Teich}(S_g))$ is also continuous, and we now prove that it is proper (note that we need to restrict the codomain to get this).

Consider a diverging sequence $m_j \in \text{Teich}(S_g)$. By compactness of $\mathbb{P}\mathscr{C}$ the sequence $[m_j] \in \mathbb{P}\mathscr{C}$ converges on a subsequence to some $[\alpha] \in$

$\mathbb{PC}$. For each j there is a $\lambda_j > 0$ such that $\lambda_j m_j \to \alpha$ in $\mathscr{C}$. Since m_j diverges in $\mathrm{Teich}(S_g)$ and hence in $\mathscr{C}$ we get $\lambda_j \to 0$. We get

$$i(\alpha, \alpha) = \lim_{j \to \infty} i(\lambda_j m_j, \lambda_j m_j) = -\pi^2 \chi(S_g) \lim_{j \to \infty} \lambda_j^2 = 0.$$

In particular $[\alpha] \notin \pi(\mathrm{Teich}(S_g))$. Therefore the restriction $\pi \colon \mathrm{Teich}(S_g) \to \pi(\mathrm{Teich}(S_g))$ is proper and hence a homeomorphism. $\qquad\square$

We will since now consider both $\mathscr{S}$ and $\mathrm{Teich}(S_g)$ embedded in $\mathbb{PC}$.

8.2.16. Thurston's compactification. We now consider the closure $\overline{\mathrm{Teich}(S_g)}$ of $\mathrm{Teich}(S_g)$ inside $\mathbb{PC}$. This closure is compact since $\mathbb{PC}$ is, and it is called the *Thurston compactification* of Teichmüller space. We define its *boundary* simply as

$$\partial \mathrm{Teich}(S_g) = \overline{\mathrm{Teich}(S_g)} \setminus \mathrm{Teich}(S_g).$$

Our aim is now to identify the topology of the Thurston boundary. The proof of Proposition 8.2.33 already shows the following.

Proposition 8.2.34. *The Thurston boundary consists of projective currents $[\alpha]$ with $i(\alpha, \alpha) = 0$.*

We will later prove that the Thurston boundary actually consists of *all* the projective currents $[\alpha]$ with $i(\alpha, \alpha) = 0$. For the moment we content ourselves with some examples.

8.2.17. Weighted pants decompositions. We construct some geodesic currents based on geodesic pants decompositions. We will need the following.

Exercise 8.2.35. Let P be a hyperbolic geodesic pair-of-pants. Every simple geodesic γ in the interior of P is open and has both ends that converge to ∂P, winding around it.

Hint. Decompose P into two right-angled hexagons. $\qquad\square$

Let a *weighted pants decomposition* for S_g be a pants decomposition $\mu = \gamma_1 \sqcup \ldots \sqcup \gamma_{3g-3}$ equipped with some real numbers $\lambda_1, \ldots, \lambda_{3g-3} \geqslant 0$. A weighted pants decomposition defines a geodesic current $\mu = \lambda_1 \gamma_1 + \ldots + \lambda_{3g-3} \gamma_{3g-3}$ such that $i(\mu, \mu) = 0$.

We will soon prove that $[\mu]$ lies in the Thurston boundary if $\mu \neq 0$. We will need the following lemma that characterises the positively weighted pants decompositions among all currents.

Lemma 8.2.36. *Let $\mu = \sum \lambda_i \gamma_i$ be a weighted pants decomposition with all positive weights $\lambda_i > 0$. A current $\alpha \in \mathscr{C}$ has $i(\alpha, \mu) = 0$ if and only if $\alpha = \sum \lambda_i' \gamma_i$ for some weights $\lambda_i' \geqslant 0$.*

Proof. If $\alpha = \sum \lambda_i' \gamma_i$, then clearly $i(\alpha, \mu) = 0$. Conversely, let α be a current with $i(\alpha, \mu) = 0$. Since $\lambda_i > 0$, the support of μ is the preimage in $\mathbb{H}^2$ of the geodesic pants decomposition. Since $i(\alpha, \mu) = 0$, the support of α consists of lines that are either contained or disjoint from that of μ.

By Exercise 8.2.35 a line l of the latter type has both its endpoints at the endpoints of some line of μ, so there are only countably many of them overall; they are not atomic since their projection in S_g is not closed and hence have zero mass. So such lines l do not occur. The support of α is hence contained in that of μ and the proof is complete. $\square$

We now characterise simple closed curves.

Lemma 8.2.37. *Let $\gamma \in \mathscr{S}$ be a simple closed curve and $\mu \in \mathscr{C}$ be a non-trivial current. If $i(\gamma, \alpha) = 0$ implies $i(\mu, \alpha) = 0$ for all $\alpha \in \mathscr{S}$, then we have $[\mu] = [\gamma]$.*

Proof. Extend $\gamma = \gamma_1$ to a geodesic pants decomposition $\beta = \sum_i \gamma_i$. The previous lemma implies that $\mu = \sum_i \lambda_i \gamma_i$ and by letting the pants decomposition vary we get $\mu = \lambda_1 \gamma_1$. $\square$

8.2.18. Pinching and twisting. How can we construct sequences of hyperbolic metrics on S_g that converge to some point at infinity? We can do this quite easily by *pinching* a simple geodesic, or by *twisting* along it.

To define the former, we need to fix some Fenchel–Nielsen coordinates (l_i, θ_i) for $\mathrm{Teich}(S_g)$ based on a pants decomposition $\mu = \gamma_1 \sqcup \ldots \sqcup \gamma_{3g-3}$. Fix a metric $(l_i, \theta_i) \in \mathrm{Teich}(S_g)$. A *pinching* along $\gamma = \gamma_1$ is any sequence of metrics where the first length coordinate l_1 tends to zero and all the other coordinates are kept bounded (from above and below).

Proposition 8.2.38. *The limit of a pinching is $[\gamma]$.*

Proof. By the Collar Lemma 7.3.7, the length of a closed geodesic α in the pinched metrics tends to infinity if $i(\gamma, \alpha) > 0$ and stays bounded if $i(\gamma, \alpha) = 0$. Therefore the pinched metrics tend in $\mathbb{P}\mathscr{C}$ to a class $[\mu]$ such that $i(\gamma, \alpha) = 0 \Rightarrow i(\mu, \alpha) = 0$ for all $\alpha \in \mathscr{S}$. Lemma 8.2.37 gives $[\mu] = [\gamma]$. $\square$

Let $m \in \mathrm{Teich}(S_g)$ be any metric and γ be a simple closed geodesic. We denoted by m_θ^γ the metric obtained from m via an earthquake of angle θ along γ, see Section 7.2.1.

Proposition 8.2.39. *The limit of m_θ^γ as $\theta \to \pm\infty$ is $[\gamma]$.*

Proof. By Proposition 7.2.10 the length of a closed geodesic α tends to infinity if $i(\gamma, \alpha) > 0$ and is constant if $i(\gamma, \alpha) = 0$. We conclude as above that the metrics tend to $[\gamma]$. $\square$

To appreciate the qualitative difference between pinching and twisting along γ, recall the explicit torus case from Section 7.2.4: if we pinch along

γ, the sequence of metrics converge to $[\gamma] \in \partial \mathbb{H}^2$ roughly like a geodesic pointing to $[\gamma]$, while if we twist along γ it does so along a horosphere centred at $[\gamma]$.

We have discovered in particular that the set $\mathscr{S}$ is entirely contained in Thurston's boundary. To fully identify the Thurston boundary, we now analyse the geodesic currents α lying in the "light cone", that is those with $i(\alpha, \alpha) = 0$. We already know that the weighted pants decompositions are there, but there is more: we will soon see that the light cone contains many new fascinating objects called *laminations*.

8.3. Laminations

In this section we characterize geometrically the currents $\alpha \in \mathscr{C}$ contained in the "light cone," *i.e.* those with $i(\alpha, \alpha) = 0$. Every such current may be represented as a *measured geodesic lamination*, a closed subset of S_g foliated by geodesics and equipped with a transverse measure. These unexpected objects play an important part in the topology of manifolds of dimension two and three.

8.3.1. Geodesic laminations. Let $S = \mathbb{H}^2/_\Gamma$ be a hyperbolic surface. Recall that in this chapter we consider geodesics only as subsets of S, neglecting their parametrisation.

A *geodesic lamination* λ is a set of disjoint simple complete geodesics in S, whose union is a closed subset of S. Each geodesic may be closed or open and is called a *leaf*; their union is the *support* of λ. We will often confuse λ with its support for simplicity, since the support determines the set of geodesics in all the interesting cases (see below).

The following examples of geodesic laminations are fundamental:

- a geodesic multicurve (that is a finite set of disjoint simple closed geodesics) in S;
- a set of disjoint lines in $\mathbb{H}^2$ whose union is closed.

A lamination in $\mathbb{H}^2$ may be particularly complicated, see Figure 8.4. Recall that $\mathscr{G} = \mathscr{G}(\mathbb{H}^2)$ is the set of all lines in $\mathbb{H}^2$, with its natural topology.

Exercise 8.3.1. A set λ of disjoint lines in $\mathbb{H}^2$ forms a closed set in $\mathbb{H}^2$ if and only if λ is closed when considered as a subset of $\mathscr{G}$.

If a set of disjoint lines in $\mathbb{H}^2$ is not closed, it suffices to take its closure to get a lamination. Every lamination in $S = \mathbb{H}^2/_\Gamma$ lifts to a Γ-invariant lamination in $\mathbb{H}^2$, hence the laminations in S are in natural bijection with the Γ-invariant laminations in $\mathbb{H}^2$.

8.3.2. Local behaviour. A geodesic lamination is often too complicated to be determined with full precision. In fact, its topology is already quite involved, and we now start by examining it locally. Let λ be a geodesic lamination on a hyperbolic surface $S = \mathbb{H}^2/_\Gamma$. The following exercise may be proved by passing to the universal cover $\mathbb{H}^2$.
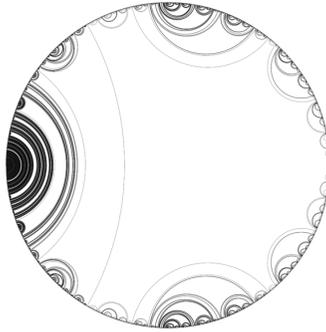
Figure 8.4. A geodesic lamination in $\mathbb{H}^2$.

Exercise 8.3.2. Every point $p \in \lambda$ has an open neighbourhood U and a chart $U \to (-1, 1) \times (-1, 1)$ that sends p to $(0, 0)$ and $\lambda \cap U$ to $(-1, 1) \times J$ for some closed subset $J \subset (0, 1)$.

What kind of $J \subset (0, 1)$ may arise? It is possible to construct some laminations in $\mathbb{H}^2$ or in some simple hyperbolic surfaces like cusps and tubes, where their support is the whole manifold. In this case J is the full interval $(0, 1)$. These cases are however of no interest for us, because of the following.

Proposition 8.3.3. *If S has finite volume, then J has empty interior.*

Proof. Suppose by contradiction that J contains some open interval. This open interval determines a set of leaves of λ, whose lift in $\mathbb{H}^2$ forms an open subset $U \subset \mathbb{H}^2$ foliated into lines. Its limit in $\partial \mathbb{H}^2$ has non-empty interior S. By Corollary 5.1.14 there is a hyperbolic transformation $\gamma \in \Gamma$ with an attracting limit point in S. This implies easily that there are γ-translates of some lines in U that intersect some lines of U transversely, a contradiction.   □

Corollary 8.3.4. *If S has finite volume, every geodesic lamination λ in S has empty interior and is determined by its support.*

8.3.3. Complementary regions. Let $\lambda \subset S$ be a geodesic lamination in a hyperbolic surface $S = \mathbb{H}^2/_\Gamma$. A *complementary region* (shortly, a *region*) is a connected component of the open complement $S \setminus \lambda$. The abstract completion of a complementary region is a hyperbolic surface with non-empty geodesic boundary consisting of lines and/or circles.

In general, a complementary region may have genus and/or infinitely many boundary components. Note that in any case there are at most countably many regions, and hence only countably many leaves of λ may be incident to a region: if λ has uncountably many leaves (this will be the

typical case), most leaves are not incident to any region (compare this with the Cantor subset in $[0, 1]$, which contains uncountably many points, but only countably many of them are adjacent to some complementary open segment).

Proposition 8.3.5. *A geodesic lamination $\lambda \subset S_g$ in a closed hyperbolic surface $S_g = \mathbb{H}^2/\Gamma$ has at most $4g - 4$ complementary regions. The boundary of each region has finitely many components.*

Proof. Every complementary region has area at least π (the area of an ideal triangle) and a region with infinitely many boundary components has infinite area. By Gauss-Bonnet we get $\mathrm{Area}(S_g) = -2\pi\chi(S_g) = (4g - 4)\pi$. □

A lamination $\lambda \subset S_g$ is *full* if every region is an ideal polygon.

8.3.4. Transverse measures. Let $\lambda \subset S$ be a geodesic lamination in a hyperbolic surface S. A *transverse arc* to λ is the support of a simple regular curve $\alpha\colon [a, b] \to S$ transverse to each leaf of λ, whose endpoints $\alpha(a)$ and $\alpha(b)$ are not contained in λ.

Definition 8.3.6. A *transverse measure* for a lamination $\lambda \subset S$ is a locally finite Borel measure L_α on each transverse arc α such that:

(1) if $\alpha' \subset \alpha$ is a sub-arc of α, the measure $L_{\alpha'}$ is the restriction of L_α;
(2) the support of L_α is $\alpha \cap \lambda$;
(3) the measure is invariant through isotopies of transverse arcs.

In particular every arc α transverse to λ has a finite *length*, defined as the total measure of the arc. The arc has length zero if and only if $\alpha \cap \lambda = \varnothing$. A *measured geodesic lamination* is a geodesic lamination equipped with a transverse measure.

Example 8.3.7. A geodesic multicurve $\lambda \subset S$ has a natural transverse measure: for any transverse arc α, the measure L_α on α is just the Dirac measure supported on the finite set $\alpha \cap \lambda$.

More generally, we may assign a positive weight $a_i > 0$ at each component λ_i of λ and define a measured geodesic lamination by giving the weight a_i at each intersection $\alpha \cap \gamma_i$. By varying weights we get distinct measured geodesic laminations with the same support.

8.3.5. Currents and measured geodesic laminations. Let $S_g = \mathbb{H}^2/\Gamma$ be a closed hyperbolic surface. We now construct a natural bijection between the measured geodesic laminations on S_g and the geodesic currents λ with $i(\lambda, \lambda) = 0$, *i.e.* those lying in the "light cone." We see a measured geodesic lamination in S_g as a Γ-invariant measured geodesic lamination in $\mathbb{H}^2$.

Let $\lambda \in \mathscr{C}$ be a geodesic current with $i(\lambda, \lambda) = 0$. By Remark 8.2.24 the support of λ is a closed Γ-invariant subset of $\mathscr{G}$ formed by disjoint

lines, hence a Γ-invariant lamination in $\mathbb{H}^2$. The geodesic current induces also a transverse measure as follows. Let α be an arc transverse to the lamination λ. Up to cutting α in finitely many arcs we may suppose that it intersects each leaf of λ in at most one point. We define the measure of a Borel set $U \subset \alpha$ as the λ-measure in $\mathscr{G}$ of the lines in λ that it intersects.

Exercise 8.3.8. This transverse measure satisfies the axioms of Definition 8.3.6 and gives λ the structure of a Γ-invariant measured geodesic lamination.

Proposition 8.3.9. *We have just defined a bijection*

$$\left\{\text{currents } \lambda \text{ with } i(\lambda, \lambda) = 0\right\} \longleftrightarrow \left\{\text{measured geodesic laminations in } S_g\right\}$$

Proof. We define the inverse map by transforming a Γ-invariant measured geodesic lamination λ into a geodesic current with support $\lambda \subset \mathscr{G}$ as follows. For every leaf l of λ, we pick a small transverse arc α that intersects all the leaves in a neighbourhood U_l of l in λ, each once. The measure on α translates into a measure on U_l, and all the measures on these small sets U_l match to yield a measure on λ thanks to Proposition 1.3.4. $\square$

We denote by $\mathscr{ML} \subset \mathscr{C}$ the set of all the measured geodesic laminations on S_g, henceforth identified with the currents lying in the "light cone". With this identification $\mathscr{ML}$ is independent of the auxiliary hyperbolic metric on S_g.

Recall that a multicurve on S_g is a finite collection of disjoint non-trivial simple closed curves. A multicurve determines a measured geodesic lamination: consider n parallel components as a single one with weight n, take the geodesic representative of the resulting essential multicurve (keeping the same weights), and use Example 8.3.7. We obtain the inclusions

$$\mathscr{S} \subset \mathscr{M} \subset \mathscr{ML} \subset \mathscr{C}$$

where $\mathscr{S}$ and $\mathscr{M}$ are the simple closed curves and multicurves in S_g.

8.3.6. Some properties. We can now use our knowledge on the geodesic currents to get some information about the measured geodesic laminations.

Proposition 8.3.10. *Let $\lambda \subset S_g$ be a measured geodesic lamination. For every $p \in \partial\mathbb{H}^2$, there are at most two lines in the preimage $\tilde{\lambda} \subset \mathbb{H}^2$ incident to p, and if there are two, none of them projects to a closed geodesic in λ.*

Proof. If there are at least three lines l_1, l_2, l_3 incident to p, with l_2 lying strictly between l_1 and l_3, we may find a small arc α transverse to $\tilde{\lambda}$, whose support lies between l_1 and l_3, that intersects l_2 and possibly other lines incident to p, and no other line of $\tilde{\lambda}$ (use Proposition 8.3.3 near l_2 here). The lines intersected by α form a pencil with positive mass. By

Proposition 8.2.8 the pencil consists of a single line, the axis of a hyperbolic element $\varphi \in \Gamma$.

On the other hand, if p is the endpoint of the axis of some φ, then no other leaf $l \subset \tilde{\lambda}$ can be incident to p, otherwise infinitely many lines $\varphi^i(l)$ would also be, and the same argument above would give a contradiction.
□

Exercise 8.3.11. There are some geodesic laminations in S_g that do not support any transverse measure.

Hint. Construct a geodesic lamination $\lambda \subset S_g$ consisting of some disjoint simple closed curves and finitely many open leaves that wind around them (see Section 7.4.5). The pre-image $\tilde{\lambda} \subset \mathbb{H}^2$ contradicts Proposition 8.3.10.
□

8.3.7. Intersection form. What is the geometric meaning of the intersection form i of two geodesic currents, when applied to a measured geodesic lamination? Let λ be a measured geodesic lamination in a hyperbolic closed surface $S_g = \mathbb{H}^2/_\Gamma$. Three cases are relevant for us:

(1) If γ is a simple closed geodesic, then either γ is contained in λ, and we get $i(\gamma, \lambda) = 0$, or γ is disjoint from λ, and we still get $i(\gamma, \lambda) = 0$, or γ is transverse to λ, and in that case $i(\lambda, \gamma)$ equals the full measure $L(\gamma)$ of γ as a transverse curve to λ.
(2) If λ' is another measured geodesic lamination, then $i(\lambda, \lambda')$ "counts" the (possibly infinitely many) transverse intersections of λ and λ' with respect to the product of the two transverse measures.
(3) If m is a hyperbolic metric, then $i(\lambda, m)$ measures the "length" of λ with respect to m, that is the integral over λ of the usual length weighted with the transverse measure.

We now consider a non-geodesic non-trivial simple closed curve γ transverse to λ, and we denote by $L(\gamma)$ its full transverse λ-measure. The quantities $L(\gamma)$ and $i(\gamma, \lambda)$ need *not* to be equal in this case (because γ is not geodesic), but this phenomenon is easily controlled by a suitable version of the bigon criterion.

Proposition 8.3.12. We have $L(\gamma) \geqslant i(\gamma, \lambda)$ and the equality holds $\Longleftrightarrow \gamma$ forms no bigon with any leaf of λ.

Proof. We follow the proof of Theorem 6.3.10. Let $\tilde{\lambda}$ be the counterimage of λ in $\mathbb{H}^2$. A lift $\tilde{\gamma}$ of γ has distinct endpoints in $\partial\mathbb{H}^2$ and $\tilde{\gamma}$ must intersect the leaves of $\tilde{\lambda}$ whose endpoints are linked with them. The curve $\tilde{\gamma}$ intersects each such leaf only once and no other leaf $\Leftrightarrow \tilde{\gamma}$ forms no bigon with any leaf of $\tilde{\lambda} \Leftrightarrow \gamma$ forms no bigon with any leaf of λ.   □

Proposition 8.3.13. If $\lambda \in \mathcal{ML}$ is full, we have $i(\lambda, \alpha) > 0$ for every current α that is not a measured sublamination of λ.
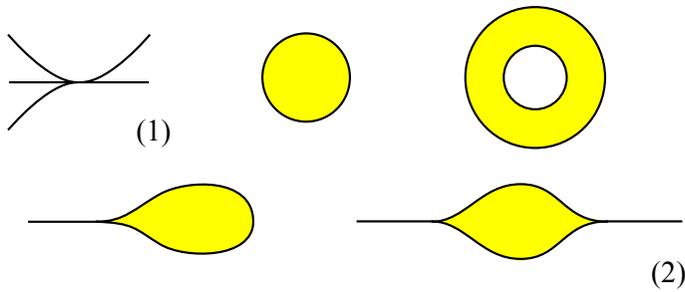
(1)

(2)

Figure 8.5. A switch contains $a \geqslant 1$ and $b \geqslant 1$ branches on each side: here $a = 3$ and $b = 2$ (1). The complementary regions of a train track contain no discs, annuli, monogons, or bigons (2).
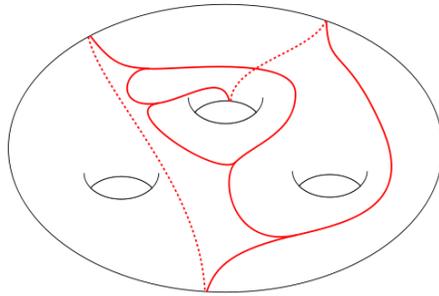

Figure 8.6. A train track on the genus-three surface S_3.

Proof. If $i(\lambda, \alpha) = 0$ then every line l in the support of α is either a leaf of λ or an infinite diagonal in some complementary ideal polygon; a diagonal would be an atomic point in α that would cover a closed geodesic and could be added to λ contradicting Proposition 8.3.10.   □

8.3.8. Train tracks. How can one construct a measured geodesic lamination, concretely? There are some nice combinatorial tools designed to this purpose, called train tracks.

A *train track* in a closed surface S_g is a closed subset $\tau \subset S_g$ built by taking a finite set of points (called *vertices* or *switches*) and joining them with disjoint arcs called *branches*. We require that every switch looks locally like Figure 8.5-(1): all branches have the same tangent line, and there is at least one branch at each side.

The *complementary regions* of τ are the abstract closures of the connected components of $S_g \setminus \tau$. Each complementary region is naturally a compact surface with finitely many singular vertices in its boundary having

"zero interior angle". In the definition of a train track, we also require that no complementary region be a disc, an annulus, a monogon, or a bigon, as in Figure 8.5. An example is shown in Figure 8.6. The exclusion of these complementary regions is essential to get the following.

Exercise 8.3.14. A train track contains at most $-6\chi(S_g)$ switches and at most $-2\chi(S_g)$ regions; it contains $-6\chi(S_g)$ switches if and only if every switch is trivalent and every complementary region is a triangle. The sphere and the torus contain no train track.

Hint. Make the appropriate Euler characteristic computation.   □

Exercise 8.3.15. Construct a train track with $-2\chi(S_g)$ triangular complementary regions in S_g for every $g \geqslant 2$.

Hint. Take a pants decomposition and add three arcs (tangent to the boundary) inside each pants to cut it into two triangular regions.   □

We obtain a finiteness property.

Corollary 8.3.16. *There are only finitely many train tracks τ in S_g up to diffeomorphism (but not up to isotopy!).*

Proof. There are only finitely many combinatorial types for τ, its complementary regions, and the way they are adjacent.   □

8.3.9. Weight systems. A *weight system* on a train track $\tau \subset S_g$ is the assignment of a non-negative real number, called *weight*, to each branch of τ, such that at every vertex the *switch condition* holds: the sum of the weights on the left branches should be equal to the sum of the weights on the right branches.

Let w be an *integral* weight system, that is one whose weights are all integers. The integral weight system w determines a closed 1-manifold in S_g as follows: replace every branch with weight a with a parallel copies of it, and pair all the branches at every switch in the obvious way (this is possible thanks to the switch condition).

Proposition 8.3.17. *The resulting closed 1-manifold is a multicurve.*

Proof. Since no complementary region of τ is a disc or monogon, the complementary regions of the closed 1-manifold have non-positive Euler characteristic (exercise).   □

Our aim now is to use the train tracks to parametrise all the multicurves in S_g. Later on, we will extend everything from multicurves to measured geodesic laminations.
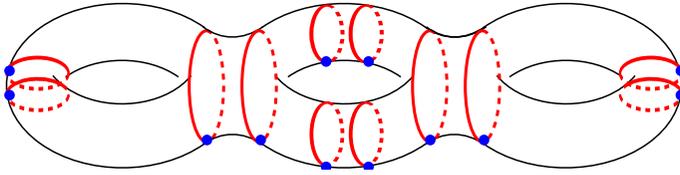
Figure 8.7. Decompose S into pairs of pants and annuli, so that each simple closed curve separates a pair of pants from an annulus. We mark a blue point in each curve.
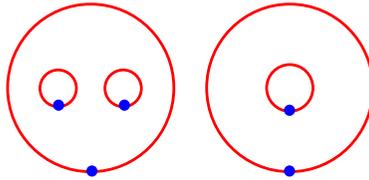

Figure 8.8. Identify each pair of pants and annulus with one of these two fixed models.

8.3.10. A parametrisation for $\mathscr{M}$. Let S_g have genus $g \geqslant 2$. We now construct finitely many train tracks that parametrise all the multicurves in S_g, and to this purpose we fix a frame similar to the one used in the Fenchel-Nielsen parametrisation of the Teichmüller space.

We decompose S_g into pairs of pants and annuli as in Figure 8.7. We mark (in blue) an arbitrary point in each closed curve and we also fix once for all a diffeomorphism (preserving orientation and marked points) between each pair of pants and annulus with the corresponding model in Figure 8.8. The resulting parametrisation of $\mathscr{M}$ will depend also on these fixed diffeomorphisms.

Let a *colouring* be the assignment of a triple a_i, b_i, c_i of non-negative numbers to each annulus of the decomposition, such that one of the following equalities holds:

$$a_i = b_i + c_i, \quad b_i = c_i + a_i, \quad c_i = a_i + b_i.$$

We assign the colour a_i to each of the two curves that bound the annulus. Since there are $-\frac{3}{2}\chi(S_g) = 3g - 3$ annuli, we get $-\frac{9}{2}\chi(S_g) = 9g - 9$ colours overall. The colouring is *integral* if the colours are integers and also the following holds: for each pair of pants, the sum $a_i + a_j + a_k$ of the colours of its boundary curves must be even.

We now transform each colouring into a weighted train track τ as follows. The boundary components of each pair-of-pants P are coloured by some triple a_i, a_j, a_k. We insert in P a portion of train track as in Figure 8.9: its shape depends on the position of the point $[a_i, a_j, a_k]$ in $\mathbb{RP}^2$ and
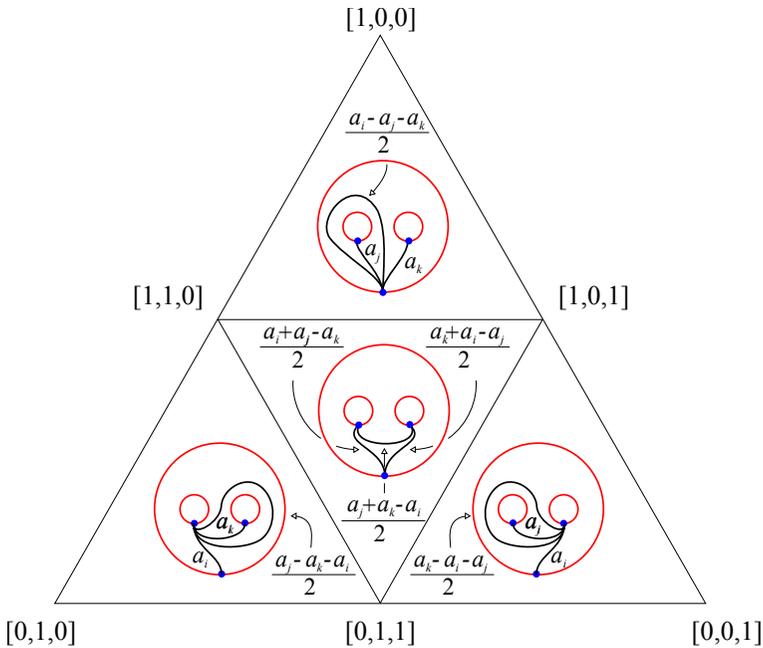
Figure 8.9. The portion of weighted train track in a pair of pants P is determined by the point $Q = [a_i, a_j, a_k] \in \mathbb{RP}^2$. Having non-negative entries, the point Q is contained in the triangle with vertices $[1, 0, 0], [0, 1, 0], [0, 0, 1]$. This triangle subdivides into four sub-triangles by adding the vertices $[0, 1, 1], [1, 0, 1], [1, 1, 0]$ and the shape of the train track depends on which sub-triangle contains Q. The three edges of the train track are given the weights indicated in the figure: these are the only weights that sum to a_i, a_j, a_k at the boundaries. When Q lies in the frontier of two or more triangles some branch has weight zero and we delete it: the different shapes prescribed by the adjacent triangles coincide after deleting this branch.

its weights depend linearly on a_i, a_j, a_k. Note that there are finitely many possible shapes overall.

We extend the train track inside each annulus A coloured with a_i, b_i, c_i as shown in Figure 8.10. Again, the shape of the portion depends on the position of $[a_i, b_i, c_i]$ in $\mathbb{RP}^2$. As a result we get a train track τ with integral weights and hence a multicurve.

Proposition 8.3.18. *The construction produces a bijection*

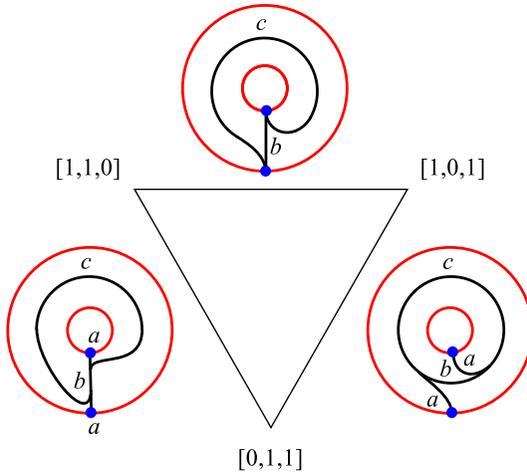$$\{\text{integral colourings}\} \longleftrightarrow \mathcal{M}.$$

Figure 8.10. The portion of weighted train track in each annulus A is determined by the point $Q = [a, b, c] = [a_i, b_i, c_i] \in \mathbb{RP}^2$. By hypothesis Q is contained in the boundary of the triangle with vertices $[1, 0, 0]$, $[0, 1, 0]$, $[0, 0, 1]$. The boundary subdivides into three segments and the shape of the train track depends on which segment contains Q. When Q is a vertex some branch has weight zero and we delete it.

Proof. We start by proving surjectivity. Given a multicurve $\mu \in \mathscr{M}$, we minimise its intersections with our decomposition into pants and annuli: now μ intersects every pair of pants P into non-trivial simple closed curves and arcs (an arc is trivial if it forms a bigon with ∂P).

By decomposing P into two hexagons, and minimising the intersections of the curves and arcs with them, one deduces easily that P contains only three isotopy classes of non-trivial closed curves (one parallel to each component of ∂P) and six isotopy classes of non-trivial arcs (one connecting every pair of components of ∂P). Note that isotopies are *not* required to fix ∂P pointwise.

We isotope all the closed curves away from P and place them inside the adjacent annuli, so there are only arcs. As above, we check easily that the isotopy class of the whole set of arcs in P is determined by the intersection numbers a_i, a_j, a_k with the components of ∂P and is hence represented by the corresponding train track portion from Figure 8.9.

We now turn to annuli. On each annulus A, there is only one isotopy class of non-trivial arcs and one isotopy class of non-trivial closed curves. However, an arc may wind many times around A and we are not allowed to use isotopies that move ∂A anymore because the configuration in P is
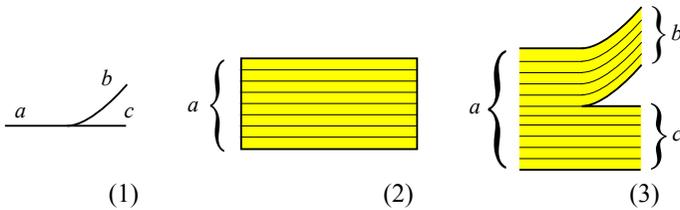
Figure 8.11. How to construct a lamination from a weighted train track. Here we have a switch with weights a, b, c (1) and the switch condition $a = b + c$ allows all the Euclidean rectangles (2) to glue (3).

already fixed: one deduces easily that the triples (a, b, c) from Figure 8.10 describe precisely all the possible configurations.

The injectivity is surprisingly simple to prove: the configuration of μ that minimises its intersections with the decomposition into pairs of pants and annuli is unique thanks to Proposition 6.3.24 and hence the numbers a_i, b_i, c_i are easily determined by μ. The proof is complete. $\square$

We have found a complete combinatorial parametrisation for the set $\mathcal{M}$, and we now want to extend it to $\mathcal{ML}$.

8.3.11. Topological laminations. A train track τ in S_g with integer weights parametrises a multicurve. We now show that, more generally, a weighted train track parametrises a measured geodesic lamination.

The construction goes as follows. First, we remove all the branches with zero weight, and we replace every branch of weight $a > 0$ with a Euclidean rectangle of width a and with arbitrary length as in Figure 8.11-(2). Thanks to the switch conditions, these rectangles glue nicely at each switch as in Figure 8.11-(3).

In Figure 8.11-(3) we see that all the lines from left- and right-rectangles are matched in a 1-1 correspondence, with a finite number of exceptions. At each exception, some $m \geqslant 1$ left-lines are matched to some $n \geqslant 1$ right-lines at some point, and we have $m + n = 3$ or 4. We call these lines and points *singular*.

After gluing the rectangles we get a closed subset λ of S_g foliated by lines, with finitely many singular points and lines. We now eliminate the singular points and lines by cutting λ carefully along them: if $m + n = 4$ this amounts simply to doubling the singular point; if $m + n = 3$ we double the singular leaf by opening a small open corridor in the foliation, starting from the singular point and digging along the leaf. Since the singular leaf may be non-compact, for the process to converge in S_g we need to shrink the width of the corridor sufficiently fast.

After this cut we get a closed subset of S_g, which we still name λ, nicely partitioned into disjoint lines that may be either open or closed:
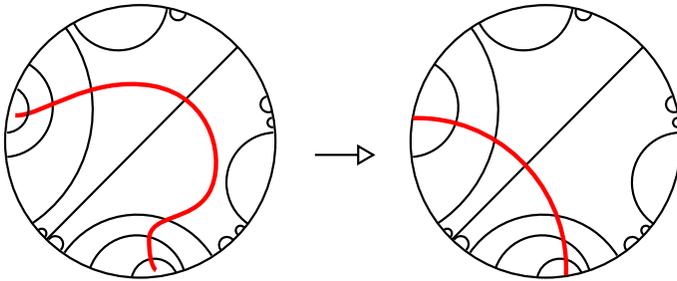
Figure 8.12. The decomposition into pants and annuli of S_g
is homotopic to a geodesic pants decomposition (with each
curve counted twice) whose counterimage in $\mathbb{H}^2$ consists of
infinitely many disjoint ultraparallel lines as in the picture.
We show that the lift of a leaf l of λ in $\mathbb{H}^2$ has two distinct
endpoints in $\partial\mathbb{H}^2$: if l is a closed curve in an annulus A,
then it is isotopic to a closed geodesic and we are done; if l
intersects annuli and pairs of pants into non-trivial arcs, its
lift intersects the ultraparallel lines forming no bigons, hence
it intersects each line at most once: therefore it has disjoint
limits at $\pm\infty$ (left). The straightening replaces the lift with
the unique line with these endpoints (right).

we call it a *topological lamination*. After the cut every rectangle of type
$L \times [0, a]$ as in Figure 8.11 transforms into a set $L \times J$ where J is obtained
by cutting $[0, a]$ along (at most) countably many points. Note that J may
be a Cantor set.

The set J inherits from $[0, a]$ a Borel measure with total mass a.
This measure gives a transverse measure to the leaves of the rectangle
$L \times J$, and λ inherits the structure of a measured topological lamination
(the notion of a transverse measure is exactly the same as in the geodesic
case).

It only remains to promote the topological lamination to a geodesic
lamination: this is usually done by *straightening* its leaves. We describe this
procedure only for a particular class of train tracks: the ones introduced
in the previous section to parametrise $\mathcal{ML}$.

8.3.12. A parametrisation for $\mathcal{ML}$. We now extend the arguments
of Section 8.3.10 from multicurves to laminations.

We fix a decomposition of S_g into pants and annuli. Every colouring
(a_i, b_i, c_i) produces a weighted train track, which in turn parametrises a
measured topological lamination λ. We fix a hyperbolic structure $S_g = \mathbb{H}^2/_\Gamma$.

The lamination λ can be *straightened* to a measured geodesic lamina-
tion $\bar{\lambda}$ as follows. A leaf l of λ is either a closed curve in some annulus A,
or it intersects every annulus A and pair of pants P in non-trivial arcs: in

both cases, every lift of l in $\mathbb{H}^2$ is a curve with two distinct limit endpoints in $\partial\mathbb{H}^2$, see Figure 8.12. We replace every lift of l with the unique line having these endpoints, and we do this for every leaf l of λ.

The result is a Γ-invariant closed set of disjoint lines in $\mathbb{H}^2$ that project to a geodesic lamination $\bar{\lambda}$ in S_g. The transverse measure on λ easily induces one on the straightened $\bar{\lambda}$: it suffices to consider transverse arcs contained in the decomposition into pairs of pants and annuli. Note that some parallel closed leaves of λ may have collapsed to a single atom closed geodesic in $\bar{\lambda}$.

Proposition 8.3.19. *The construction induces a bijection*

$$\{\text{colourings}\} \longleftrightarrow \mathscr{ML}.$$

Proof. We adapt the proof of Proposition 8.3.18, starting with surjectivity: given a measured geodesic lamination λ, we determine a colour (a_i, b_i, c_i) representing it.

The decomposition into pairs of pants and annuli is homotopic to a geodesic pants decomposition $\mu = \gamma_1 \sqcup \ldots \sqcup \gamma_{3g-3}$ (with each curve counted twice). We set $a_i = i(\gamma_i, \lambda)$.

The measured geodesic lamination $\lambda \in \mathscr{ML}$ decomposes as $\lambda = \lambda_0 \sqcup \lambda_1$ where $\lambda_0 = \sum_i k_i \gamma_i$ is a weighted pants decomposition with support in μ and λ_1 is transverse to μ. If $k_i \neq 0$ then necessarily $a_i = 0$ and we set $b_i = c_i = k_i$.

At every pair of pants P the three colours a_i, a_j, a_k determine a portion of weighted train track as prescribed by Figure 8.9. The intersection $\lambda_1 \cap P$ consists of geodesic and hence non-trivial arcs: therefore the weighted train track describes faithfully this portion of measured geodesic lamination, up to isotopy. The numbers b_i and c_i are then determined by the way these two portions wind and match along γ_i. The colours (a_i, b_i, c_i) parametrise λ.

The choices of the colours (a_i, b_i, c_i) were forced by the intersection of λ with the two pants P adjacent to γ_i (exercise). This shows injectivity. $\square$

We have parametrised $\mathscr{ML}$, and we now investigate its topology. The space of all colourings (a_i, b_i, c_i) forms the subset $C \times \ldots \times C \subset \mathbb{R}^3 \times \ldots \times \mathbb{R}^3$ where $C \subset \mathbb{R}^3$ is the cone based on the origin over the sides of the triangle with vertices $(1, 1, 0)$, $(0, 1, 1)$, and $(1, 0, 1)$, and is homeomorphic to $\mathbb{R}^2$. The space of colourings is homeomorphic to $\mathbb{R}^{2\cdot(3g-3)} = \mathbb{R}^{6g-6}$, that is to the Teichmüller space itself!

The parametrisation identifies $\mathscr{ML}$ with $C \times \ldots \times C$. We want to show that this identification is a homeomorphism, and to this purpose we study the intersection form.
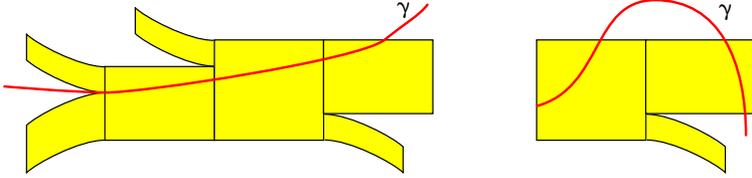
Figure 8.13. A monotonic curve γ intersects the products $L \times J$ in monotonic paths (left) and forms no bigons outside the products (right).

8.3.13. Intersection form. We now show that the intersection form i between curves and laminations has a surprisingly simple behaviour after that we parametrise the space $\mathcal{ML}$ as $C \times \ldots \times C$.

We note that $C \times \ldots \times C$ is a piecewise-linear object in $\mathbb{R}^{9g-9}$, that is it is the support of a simplicial complex. Recall that a continuous map between simplicial complexes is *simplicial* if it sends each simplex onto a simplex in an affine linear way, and it is *piecewise linear* if it restricts to a simplicial map on some subdivisions. Finally, the map is *half-integral piecewise linear* if all the affine linear maps have half-integer (possibly integer) coefficients.

Proposition 8.3.20. *For every $\gamma \in \mathscr{S}$ the map*

$$C \times \ldots \times C \to \mathbb{R}$$
$$\lambda \longmapsto i(\gamma, \lambda)$$

is half-integral piecewise linear.

Proof. A colour $(a_i, b_i, c_i) \in C \times \ldots \times C$ defines a weighted train track τ and hence a measured topological lamination λ, obtained by substituting every edge of τ with some product $L \times J$ of horizontal leaves. Thanks to Proposition 8.3.12 the intersection $i(\gamma, \lambda)$ is realized by some representative γ which is either a leaf of λ or is transverse to λ and forms no bigon with any leaf of λ.

We say that a closed curve transverse to λ is *monotonic* if

(1) it intersects every product $L \times J$ into arcs that are monotonic in both coordinates, and the same monotonicity is preserved when the curve goes from one product $L \times J$ to an adjacent one as in Figure 8.13-(left);

(2) it makes no bigons outside the rectangles as in Figure 8.13-(right).

Concerning (1), a horizontal path in $L \times J$ disjoint from λ is allowed, but it must keep being horizontal on the adjacent products.
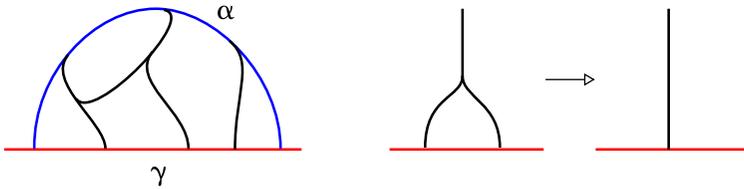
Figure 8.14. If a monotonic curve γ forms a bigon with a
leaf α, by cutting carefully τ along γ we construct a portion
of train track on this bigon containing α and with truncated
edges exiting orthogonally from γ (left). We then simplify
the triangles adjacent to γ (right).

A simple closed curve that forms no bigons with any leaf of λ can
be easily isotoped to be monotonic. Conversely, we now prove that a
monotonic curve γ forms no bigons with any leaf of λ. If it did, by cutting
τ along γ we would get an abstract bigon with a portion of train track as
in Figure 8.14-(left). If γ is adjacent to some triangles as in Figure 8.14-
(right) we close them as shown there. Finally, by doubling the bigon along
γ we get a train track on a disc, and by doubling again we build a train
track in a sphere, which is absurd (one checks easily that no complementary
region is a disc, annulus, monogon, or bigon: there are no bigons because
we have closed the triangles as in Figure 8.14-(right)).

The λ-transversal length $L(\gamma)$ of a monotonic γ is the sum of the
lengths of the sub-paths intersecting sequences of products $L \times J$ as in
Figure 8.13-(left). One checks easily that the length of each sub-path
is a half-integer linear combination of the colours (a_i, b_i, c_i). Therefore
$L(\gamma) = i(\gamma, \lambda)$ is a half-integer combination of the colours.

If we vary the colouring (a_i, b_i, c_i) a little, the curve γ keeps being
monotonic in the same way as before, except when a portion of γ is hor-
izontal: in that case the new monotonicity depends on how the colouring
varies. In all cases, the new γ is still monotonic, hence bigonless, hence
$L(\gamma) = i(\gamma, \lambda)$ again.

If there are no horizontal portions in γ, the length $L(\gamma)$ varies lin-
early with the same integral formula found above. If there are horizontal
portions, then the colour lies in a hyperplane of $\mathbb{R}^{9g-9}$, and there are two
different linear formulas joining there. Hence $i(\gamma, \lambda)$ is half-integral piece-
wise linear. $\square$

The map $\lambda \mapsto i(\gamma, \lambda)$ is also obviously homogeneous, in the sense
that $i(\gamma, t\lambda) = ti(\gamma, \lambda)$ for all $t \geqslant 0$.

The proof of Proposition 8.3.20 contains a recipe for calculating $i(\gamma, \lambda)$. It suffices to put γ in monotonic position and then add the contributions of each monotonic arc. For instance, the i-th curve γ_i of the pants decomposition has an obvious monotone position giving $i(\gamma_i, \lambda) = a_i$.

The colours b_i and c_i do not have such an immediate description, but the following holds anyway.

Exercise 8.3.21. There exist finitely many simple closed curves γ_j whose intersections $i(\lambda, \gamma_j)$ determine the colouring (a_i, b_i, c_i) representing λ in a half-integral piecewise-linear continuous fashion.

Hint. Choose some additional curves as in Proposition 7.3.4 and compute their intersections with λ.   □

8.3.14. The Thurston boundary. We can finally determine the topology of the measured geodesic laminations space $\mathscr{ML}$ and of the Thurston boundary of the Teichmüller space.

Proposition 8.3.22. *The colouring parametrisation induces a homeomorphism $\mathscr{ML} \cong \mathbb{R}^{6g-6}$.*

Proof. The map $\mathscr{ML} \to C \times \ldots \times C \cong \mathbb{R}^{6g-6}$ is continuous by Exercise 8.3.21. It is proper: if the colours (a_i, b_i, c_i) stay bounded, the intersections $i(\lambda, \gamma_i)$ with finitely many filling curves γ_i stay bounded, and hence λ moves in a compact set by Proposition 8.2.26. The continuous map is a homeomorphism by Proposition 7.3.11.   □

Let now $\mathbb{P}\mathscr{ML} \subset \mathbb{P}\mathscr{C}$ be the set of all *projective measured laminations*, that is the image of $\mathscr{ML} \setminus 0$ in $\mathbb{P}\mathscr{C}$. The set $\mathbb{P}\mathscr{ML}$ contains the projective simple closed curves $\mathbb{P}\mathscr{S}$ and the projective multicurves $\mathbb{P}\mathscr{M}$.

Theorem 8.3.23. *The following homeomorphism holds*

$$\partial \mathrm{Teich}(S_g) = \mathbb{P}\mathscr{ML} \cong S^{6g-7}.$$

The set $\mathbb{P}\mathscr{S}$ is dense in $\mathbb{P}\mathscr{ML}$.

Proof. The homeomorphism $\mathscr{ML} \cong \mathbb{R}^{6g-6}$ induces $\mathbb{P}\mathscr{ML} \cong S^{6g-7}$. Concerning the Thurston boundary, we know from Section 8.2.18 and Proposition 8.2.34 that

$$\mathbb{P}\mathscr{S} \subset \partial \mathrm{Teich}(S_g) \subset \mathbb{P}\mathscr{ML}.$$

We now show that $\mathbb{P}\mathscr{S}$ is dense in $\mathbb{P}\mathscr{ML}$, and this concludes the proof.

First, we prove that $\mathbb{P}\mathscr{M}$ is dense in $\mathbb{P}\mathscr{ML}$. Rational colours (a_i, b_i, c_i) form a dense subset of $C \times \ldots \times C$ and project to a dense subset of $\mathbb{P}\mathscr{ML}$. Every rational colour is a multiple of an integer colour, which represents a multicurve.

Second, we prove that $\mathbb{P}\mathscr{S}$ is dense in $\mathbb{P}\mathscr{M}$. Let μ be a multicurve in S_g. Up to acting via $\mathrm{MCG}(S_g)$ we may suppose that μ is supported on the pants decomposition $\gamma_1, \ldots, \gamma_{3g-3}$ used to define $C \times \ldots \times C$, hence

μ is defined by some integral colour $(a_i, b_i, c_i) = (0, b_i, b_i)$. Now pick any simple closed curve γ with $a_i' = i(\gamma, \gamma_i) > 0$ for all i, represented by some colours (a_i', b_i', c_i'). By Dehn twisting γ along the γ_i we vary the pair b_i', c_i' as we please, and by twisting at the correct rates we construct a sequence of simple closed curves that converge projectively to $[\mu]$. $\qquad\square$

The mapping class group $\mathrm{MCG}(S_g)$ acts naturally on the whole set $\mathscr{C}$ of currents and in particular it acts by homeomorphisms on $\partial\mathrm{Teich}(S_g)$.

The identification of $\partial\mathrm{Teich}(S_g) = \mathbb{P}\mathscr{ML}$ with the projectivisation of $C\times\ldots\times C$ is of course not canonical because it depends on a decomposition of S_g into pairs of pants and annuli. However, Proposition 8.3.20 and Exercise 8.3.21 imply that any two different identifications differ by some projective integral piecewise-linear homeomorphisms, hence the Thurston boundary has a natural *projective integral piecewise linear* structure, called PIP for short, which is preserved by $\mathrm{MCG}(S_g)$. In particular $\mathbb{P}\mathscr{ML}$ contains some natural rational points, and these are $\mathbb{P}\mathscr{M}$.

8.3.15. A projection. The reader has probably noted that the coordinates that yield the homeomorphisms $\mathrm{Teich}(S_g) \cong \mathbb{R}^{6g-6}$ and $\partial\mathrm{Teich}(S_g) \cong S^{6g-7}$ are quite similar: they both depend on a pants decomposition μ plus some additional marking, and every curve of the pants decomposition contributes roughly with two parameters, a "length" and a "twist". It would now be reasonable to expect that both these coordinates merge nicely to give a global homeomorphism $\overline{\mathrm{Teich}(S_g)} \cong D^{6g-6}$, but this is unfortunately not the case.

The proof that $\overline{\mathrm{Teich}(S_g)}$ is homeomorphic to D^{6g-6} is disappointingly indirect. We first prove that $\overline{\mathrm{Teich}(S_g)}$ is a topological manifold with boundary, by constructing some charts. To construct a chart, we first build a map

$$q\colon \mathrm{Teich}(S_g) \to \mathscr{ML}$$

that depends only on a fixed pants decomposition μ. For a given hyperbolic metric $m \in \mathrm{Teich}(S_g)$, we construct a measured geodesic lamination $q(m)$ as follows. We straighten μ to its geodesic representative (with respect to m), and we consider separately each geodesic pair of pants P of the decomposition. We first show that P has a natural partial foliation, that depends only on its metric, whose leaves are not geodesics. We then glue these partial foliations to get a partial foliation on S_g, and then straighten it to a geodesic lamination $q(m)$.

The natural partial foliation on P is constructed as follows. Let g_{ij} be the unique orthogeodesic connecting the i-th and j-th boundary component of P, for all $1 \leqslant i, j \leqslant 3$. Let a_1, a_2, a_3 be the lengths of the boundaries of P. We consider the point $Q = [a_1, a_2, a_3]$ in the triangle of Figure 8.9, pick the three orthogeodesics g_{ij} isotopic to the three curves indicated in the sub-triangle of Figure 8.9 containing Q, and thicken each g_{ij} to a metric R-neighbourhood of g_{ij}, where R is the linear combination of
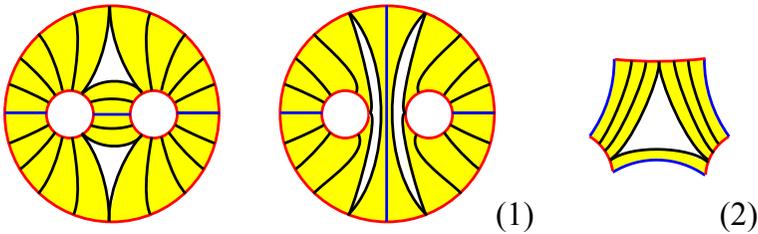
(1) (2)

Figure 8.15. Every geodesic pair of pants P has a natural partial foliation: the shape of the foliation is determined by the sub-triangle containing $Q = [a_1, a_2, a_3]$ in Figure 8.9, where a_1, a_2, a_3 are the lengths of ∂P; the two pictures shown here correspond to two cases. In each case, there are three rectangles: each is the R-neighbourhood of the corresponding (unique) orthogeodesic (drawn in blue), and R is the linear combination of a_1, a_2, a_3 prescribed by Figure 8.9, that is the unique combination that guarantees that the three rectangles match nicely as in the figure (1). We cut the geodesic pants P along the three blue orthogeodesics and get two right-angled hexagons. Since by hypothesis the three red sides have length $\geqslant \varepsilon/2$, one sees easily that the central triangle has diameter bounded by some constant C' depending only on ε, and hence every leaf in each of the three rectangles is shorter than some C'' that also depends only on ε (2).

a_1, a_2, a_3 indicated in Figure 8.9. The R-neighbourhood is naturally foliated into (non-geodesic) arcs staying at fixed distance from g_{ij} and the foliated neighbourhoods cover nicely much of the pair of pants P, as shown in Figure 8.15-(1). Every foliated rectangle is equipped with a natural transverse measure induced by the orthogonal distance between leaves.

The partial foliations of all the pairs of pants of S_g glue to a singular measured foliation for S_g, which straightens as prescribed in Section 8.3.12 to a measured geodesic lamination $q(m)$.

Let $\mu = \gamma_1 \sqcup \ldots \sqcup \gamma_{3g-3}$ be our original pants decomposition.

Proposition 8.3.24. *The map q restricts to a homeomorphism*

$$q \colon \text{Teich}(S_g) \longrightarrow \left\{ \lambda \in \mathscr{ML} \mid i(\lambda, \gamma_i) > 0 \ \forall i \right\}.$$

Proof. Compare the coordinates (l_i, θ_i) for $\text{Teich}(S_g)$ and (a_i, b_i, c_i) for $\mathscr{ML}$. The map q sends (l_i, θ_i) to (l_i, b_i, c_i) for some (b_i, c_i) that depends homeomorphically on $\theta_i \in \mathbb{R}$ for each i. Note that $a_i = i(\lambda, \gamma_i)$.
□

8.3.16. The fundamental lemma. We now show that the projection q distorts very little the lengths of the simple closed curves, as long as we put a lower bound on the lengths of the curves $\gamma_1, \ldots \gamma_{3g-3}$ of the fixed

pants decomposition μ. This technical fact is called the Fundamental Lemma [**19**].

For every $\varepsilon > 0$, we define $V(\varepsilon) \subset \text{Teich}(S_g)$ to be the open subset consisting of all metrics m such that $i(m, \gamma_i) > \varepsilon$ for all i.

Lemma 8.3.25. *For every simple closed curve $\alpha \in \mathscr{S}$ there exists a constant $C > 0$ such that, for all $m \in V(\varepsilon)$, we have*

$$i(q(m), \alpha) \leqslant i(m, \alpha) \leqslant i(q(m), \alpha) + C.$$

Proof. We represent $q(m)$ as a singular partial foliation, without any straightening. The transverse measure of $q(m)$ is just the length of orthogonal geodesics, hence the $q(m)$-measure of any piecewise-transverse closed curve is smaller or equal than its length: this proves that $i(q(m), \alpha) \leqslant i(m, \alpha)$.

We prove the other inequality. If $\alpha = \gamma_i$ then $i(m, \alpha) = i(q(m), \alpha)$, so we suppose that $i(\alpha, \mu) > 0$. Up to isotopy we may take α to be monotonically transverse to $q(m)$, recall the proof of Proposition 8.3.20. The curve α intersects each pair of pants P into some essential arcs β that cross each foliated rectangle monotonically.

We can easily homotope each of these arcs $\beta \subset \alpha \cap P$ with fixed endpoints into a (not necessarily simple!) piecewise smooth arc that decomposes into finitely many sub-arcs, that are alternatively contained either in a component of ∂P or in a leaf of some foliated rectangle. Each sub-arc in a component of ∂P may make many full turns (and hence may not be injective). Each of the three rectangles contains at most one leaf that is a sub-arc of β. We arrange the homotopy efficiently so that the new β has the same transverse $q(m)$-measure as before.

The transverse $q(m)$-measure of β on the sub-arcs in ∂P is equal to its length, whereas on the leaf sub-arcs it is zero. Therefore the m-length of β is equal to its measure plus the length of at most 3 leaves of some rectangles in P. Figure 8.15-(2) shows that each such leaf is shorter than some constant C'' that depends only on ε.

Note that α is decomposed into $i(\alpha, \mu)$ arcs like β. By homotoping each β as above we find a homotopic representative for α whose length is at most the $q(m)$-measure of α plus $C = 3i(\alpha, \mu)C''$. The length of any homotopic representative is greater or equal than the length $i(m, \alpha)$ of the geodesic one, and this proves the second inequality. $\square$

Corollary 8.3.26. *Let $m_i \in V(\varepsilon)$ be a diverging sequence in $\text{Teich}(S_g)$. The sequence converges in $\overline{\text{Teich}(S_g)} \iff$ the sequence $[q(m_i)] \in \mathbb{PML}$ does, and in this case they tend to the same limit.*

Proof. If m_i converges to some $[\alpha] \in \mathbb{PML}$, we have $\lambda_i m_i \to \alpha$ for some real numbers $\lambda_i \to 0$. By the fundamental lemma, for every simple closed curve $\gamma \in \mathscr{S}$ we get

$$\left| i(\lambda_i m_i, \gamma) - i(\lambda_i q(m_i), \gamma) \right| \to 0.$$

Therefore $\lambda_i q(m_i)$ converges (on a subsequence) to a $\beta \in \mathcal{ML}$ such that $i(\alpha, \gamma) = i(\beta, \gamma)$ for all $\gamma \in \mathcal{S}$. Exercise 8.3.21 gives $\beta = \alpha$.

The other implication is analogous.   □

8.3.17. A topological chart. Pick an arbitrary element $[\lambda] \in \mathbb{P}\mathcal{ML} = \partial \mathrm{Teich}(S_g)$. We construct an explicit neighbourhood of $[\lambda]$ in $\overline{\mathrm{Teich}(S_g)}$.

Let $\mu = \gamma_1 \sqcup \ldots \sqcup \gamma_{3g-3}$ be a pants decomposition such that $i(\gamma_i, \lambda) > 0$ for all i. We fix $\varepsilon > 0$ and define as above $V(\varepsilon) \subset \mathrm{Teich}(S_g)$ as the set of all the metrics m such that $i(\gamma_i, m) > \varepsilon$ for all i. We define similarly $W \subset \mathbb{P}\mathcal{ML}$ as the set of all $[\alpha]$ such that $i(\gamma_i, \alpha) > 0$ for all i. We have $[\lambda] \in W$.

Proposition 8.3.27. *We have*

$$\pi(q(V(\varepsilon))) = \pi(q(\mathrm{Teich}(S_g))) = W.$$

The set $W \cup V(\varepsilon)$ is an open neighbourhood of $[\lambda]$ in $\overline{\mathrm{Teich}(S_g)}$.

Proof. Proposition 8.3.24 implies the second equality, the first holds because every $[\alpha] \in W$ is represented by an α with $i(\gamma_i, \alpha) \geqslant \varepsilon$.

The sets W and $V(\varepsilon)$ are open in $\mathbb{P}\mathcal{ML}$ and $\mathrm{Teich}(S_g)$ respectively. If $W \cup V(\varepsilon)$ were not open in $\overline{\mathrm{Teich}(S_g)}$, there would be a sequence $m_i \notin V(\varepsilon)$ of metrics converging to some $[\alpha] \in W$. On a subsequence we may suppose that $i(m_i, \gamma_j) < \varepsilon$ for some fixed j and therefore $i(\alpha, \gamma_i) = 0$, a contradiction.   □

We have found an open neighbourhood of $[\lambda]$ and we now want to determine its topology.

Proposition 8.3.28. *There is a homeomorphism*

$$\phi \colon W \cup V(\varepsilon) \longrightarrow U$$

onto an open subset $U \subset H$ of a half-space $H \subset \mathbb{R}^{6g-6}$, with $\phi(W) = U \cap \partial H$.

Proof. We complete the pants decomposition μ to a filling set $\mu' \in \mathscr{C}$ of simple closed curves in S_g, considered as a current. We know from Proposition 8.2.28 that the metrics m with $i(m, \mu') \leqslant M$ form a compact subset in $\mathrm{Teich}(S_g)$. We define the map

$$\phi \colon W \cup V(\varepsilon) \longrightarrow W \times [0, 1]$$
$$x \longmapsto \begin{cases} (x, 0) & \text{if } x \in W, \\ (\pi(q(x)), e^{-i(q(x), \mu')}) & \text{if } x \in V(\varepsilon). \end{cases}$$

The map ϕ is continuous: if a sequence $m_i \in V(\varepsilon)$ of metrics tends to $[\alpha] \in \mathbb{P}\mathcal{ML}$, then $i(m_i, \mu') \to \infty$, hence $i(q(m_i), \mu') \to \infty$ by the fundamental lemma. Moreover $\pi(q(m_i)) \to [\alpha]$ by Corollary 8.3.26. Summing up, we get that $\phi(m_i) \to ([\alpha], 0) = \phi([\alpha])$.

The map ϕ is injective: given two metrics m, m', if the first components of $\phi(m), \phi(m')$ are equal then $q(m) = \lambda q(m')$, and if the second components are equal we get $\lambda = 1$.

By similar methods one proves that the image of ϕ is open and the inverse there is continuous. The set $W \times [0, 1]$ embeds in H since W embeds in $\mathbb{R}^{6g-7}$. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Corollary 8.3.29. *The space* $\overline{\mathrm{Teich}(S_g)}$ *is a topological manifold with boundary* $\partial\mathrm{Teich}(S_g)$.

8.3.18. Conclusion. We have discovered that $\overline{\mathrm{Teich}(S_g)}$ is a compact topological manifold with boundary, and we now invoke Theorem 1.7.18 to deduce that it is a disc. We have finally completed the proof of the following theorem, which is the main achievement of the whole chapter.

Theorem 8.3.30. *The closure* $\overline{\mathrm{Teich}(S_g)}$ *of* $\mathrm{Teich}(S_g)$ *in* $\mathbb{P}\mathscr{C}$ *is homeomorphic to the closed disc* D^{6g-6}. *Its interior is* $\mathrm{Teich}(S_g)$ *and its boundary contains* $\mathscr{S}$ *as a dense subset.*

Thurston's original compactification theorem embeds everything in $\mathbb{R}^{\mathscr{S}}$ instead of $\mathscr{C}$. We can easily deduce it from Theorem 8.3.30.

Proof of Theorem 8.1.5. The natural map $\mathscr{C} \to \mathbb{R}^{\mathscr{S}}$ induced by the intersection form i is continuous since i is, and it induces a continuous map $\varphi\colon \mathbb{P}\mathscr{C} \to \mathbb{P}(\mathbb{R}^{\mathscr{S}})$ on their projective spaces.

The Teichmüller space is embedded in both projective spaces and φ restricts to a homeomorphism between the two embeddings, and to a continuous surjective map from their compactifications. The map on compactifications is actually injective since $\varphi|_{\mathbb{P}\mathscr{ML}}$ is (intersections with simple closed curves distinguish laminations) and is hence a homeomorphism. $\qquad\square$

8.4. Surface diffeomorphisms

We have discovered that Thurston's compactification $\overline{\mathrm{Teich}(S_g)}$ of the Teichmüller space is homeomorphic to a closed disc. The mapping class group $\mathrm{MCG}(S_g)$ acts on it naturally. We can now apply Brouwer's fixed point theorem to every element $\varphi \in \mathrm{MCG}(S_g)$, and characterise φ according to the position of its fixed points, much similarly as we did for the isometries of $\mathbb{H}^n$.

8.4.1. The torus case. As usual, the flat torus case is very instructive because everything can be written explicitly.

We know from Proposition 7.1.5 that $\mathrm{MCG}(T)$ acts on $\mathrm{Teich}(T)$ like the Möbius transformations $\mathbb{PSL}_2(\mathbb{Z})$ do on the hyperbolic half-plane H^2. The action of course extends to the compactification $\overline{H^2} = H^2 \cup \mathbb{R} \cup \{\infty\}$ and every non-trivial isometry $A \in \mathbb{PSL}_2(\mathbb{Z})$ is elliptic, parabolic, or hyperbolic according to the position of its fixed points.

As an integral matrix, the isometry A has also more properties that are easy to check: if it is elliptic, it has finite order because $\mathbb{PSL}_2(\mathbb{Z})$ is

discrete; if it is parabolic, it is conjugate in $\mathbb{PSL}_2(\mathbb{Z})$ to a matrix $\left(\begin{smallmatrix} 1 & n \\ 0 & 1 \end{smallmatrix}\right)$ for some $n \neq 0$ and is thus the n-th power of a Dehn twist; if it is hyperbolic, it is conjugate in $\mathbb{PSL}_2(\mathbb{R})$ to a diagonal matrix $\left(\begin{smallmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{smallmatrix}\right)$ for some $\lambda > 1$, with some basis of eigenvectors v, w. The two foliations of $\mathbb{R}^2$ into lines parallel to v or w are both preserved by A and descend to foliations in $T = \mathbb{R}^2 / \mathbb{Z}^2$ that are preserved by A. The two foliations are both irrational (because λ is), that is every leaf is dense in T, and A stretches one foliation by λ and contracts the other by $1/\lambda$. One such diffeomorphism of T is called *Anosov*.

Summing up, the non-trivial elements in $\mathrm{MCG}(T)$ either have finite order, or preserve a simple closed curve, or are Anosov. We now define an analogous trichotomy for the elements in $\mathrm{MCG}(S_g)$ when $g \geqslant 2$, where the foliations are replaced by measured geodesic laminations, and the measure is there to encode stretchings and contractions.

8.4.2. The trichotomy. Let S_g have genus $g \geqslant 2$. The mapping class group $\mathrm{MCG}(S_g)$ acts naturally on the whole space $\mathscr{C}$ of currents and in particular on the compactification $\overline{\mathrm{Teich}(S_g)} \cong D^{6g-6}$ of the Teichmüller space.

Let $\varphi \in \mathrm{MCG}(S_g)$ be a non-trivial element. By Brouwer's fixed point theorem, φ fixes at least one point in $\overline{\mathrm{Teich}(S_g)}$. We say that φ is:

(1) *finite order* if it fixes a hyperbolic metric $m \in \mathrm{Teich}(S_g)$;
(2) *reducible* if it fixes a multicurve $\mu \in \mathscr{M}$;
(3) *pseudo-Anosov* in all the other cases.

We now analyse the three cases individually.

8.4.3. Finite order elements. We must first explain the terminology.

Proposition 8.4.1. *A non-trivial element $\varphi \in \mathrm{MCG}(S_g)$ is finite order if and only if it has indeed finite order in $\mathrm{MCG}(S_g)$.*

Proof. Suppose that φ preserves the isotopy class $[m] \in \mathrm{Teich}(S_g)$ of a hyperbolic metric m in S_g. We can choose a representative for φ that fixes m. This representative is an isometry for m. Since the isometry group of a closed hyperbolic manifold is finite (see Corollary 4.3.8) we have $\varphi^n = \mathrm{id}$ for some $n > 1$ and φ has indeed finite order in $\mathrm{MCG}(S_g)$.

Conversely, let φ be an element having finite order in $\mathrm{MCG}(S_g)$. The subgroup $\langle \varphi \rangle$ generated by φ cannot act freely on $\mathrm{Teich}(S_g) \cong \mathbb{R}^{6g-6}$, otherwise it would quotient $\mathbb{R}^{6g-6}$ to an aspherical manifold with finite fundamental group, contradicting Theorem 1.7.10. Therefore some non-trivial power of φ has a fixed point in $\mathrm{Teich}(S_g)$.

If φ has prime order we easily conclude that also φ has a fixed point and we are done. However, if φ has order $p_1 \cdots p_s$ for some primes p_i, we need to do more work. By induction, $\varphi' = \varphi^{p_1 \cdots p_{s-1}}$ has a fixed point $[m] \in \mathrm{Teich}(S_g)$ and is hence represented by an isometry for S_g with metric m. The isometry φ' quotients S_g to a hyperbolic orbifold, and the fixed points

Fix(φ') of φ' in Teich(S_g) can be identified naturally with the (suitably defined) Teichmüller space of this orbifold, which is homeomorphic to $\mathbb{R}^N$ for some $N > 0$ like in the surface case (exercise).

Since φ and φ' commute, the first act as a mapping class on Fix(φ') with order $p_1 \cdots p_{s-1}$: we conclude by induction on s that φ has a fixed point in Teich(S_g). $\qquad\square$

We get in particular the following corollary, which is far from obvious because MCG(S_g) does *not* act itself on S_g, and algebraic relations in MCG(S_g) do not translate into algebraic relations between representatives, except in some very lucky cases.

Corollary 8.4.2. *If $\varphi \in$ MCG(S_g) has order k, it may be represented by a diffeomorphism $\varphi\colon S_g \to S_g$ such that $\varphi^k = $ id.*

Proof. The class φ has a representative $\varphi\colon S_g \to S_g$ that is an isometry for some hyperbolic metric; the isometry φ^k is isotopic to the identity and is hence the identity by Corollary 4.3.7. $\qquad\square$

8.4.4. Reducible elements. We must explain the terminology also in this case. If φ fixes a multicurve μ, one can cut S_g along μ and look at the restriction of φ to the resulting pieces: after extending all the theory to surfaces with boundary (that we have not done here for simplicity), we can hence study inductively each piece, and this explains the word *reducible*.

The cases (1) and (2) are not exclusive: there are isometries of hyperbolic surfaces that preserve some multicurve. On the other hand, there are finite order elements that are not reducible (exercise) and reducible mapping classes that are not of finite order: for instance, Dehn twists.

8.4.5. The action on $\mathscr{ML}$. The mapping class group MCG(S_g) of S_g acts on the currents space $\mathscr{C}$ and hence on the "light cone" subspace $\mathscr{ML}$ of all measured geodesic laminations, which contains the space $\mathscr{M}$ of multicurves.

The action of MCG(S_g) on $\mathscr{ML}$ can be seen concretely inside S_g. It suffices to consider $\mathscr{ML}$ as the space of all measured *topological* laminations (defined in Section 8.3.11) considered up to isotopy and collapsing of parallel closed leaves. Now $\varphi \in$ MCG(S_g) acts on $\mathscr{ML}$ simply by sending the measured topological lamination μ to $\varphi(\mu)$.

We now prove that if φ fixes a non-trivial point in $\mathscr{ML}$ we fall back into one of the two cases already considered.

Proposition 8.4.3. *If $\varphi(\mu) = \mu$ for some non-trivial $\mu \in \mathscr{ML}$, then φ is either finite order or reducible.*

Proof. We fix a hyperbolic metric on S_g and represent μ as a measured geodesic lamination there.

By Proposition 8.3.5 there are finitely many complementary regions in $S_g \setminus \mu$. If μ is not full, some region is not an ideal polygon and hence

deformation-retracts onto a subsurface $S' \subset S_g$, whose boundary components $\partial S'$ are non-trivial simple closed curves in S_g. The union of all such curves $\partial S'$ as S' varies produces a multicurve preserved by φ, which is hence reducible.

If μ is full, all the complementary regions are ideal polygons, and after substituting φ with a finite power we can suppose that each region is fixed (not pointwise) by φ, together with each of its boundary components. We consider the preimage $\tilde{\mu} \subset \mathbb{H}^2$ and note that its complementary regions in $\mathbb{H}^2$ are still ideal finite polygons, the lifts of the ones in S_g.

The diffeomorphism φ lifts to a homeomorphism $\tilde{\varphi} \colon \overline{\mathbb{H}^2} \to \overline{\mathbb{H}^2}$ that fixes $\tilde{\mu}$, and we may suppose that it fixes (not pointwise) a complementary polygon and each of its boundary lines. In particular φ fixes orientation-preservingly a line $l \subset \tilde{\mu}$.

Remember that $\mathcal{G}$ is the set of all lines in $\mathbb{H}^2$. We now prove that, if $\tilde{\varphi}$ fixes (orientation-preservingly) a line $l \subset \tilde{\mu} \subset \mathcal{G}$, then it fixes (still orientation-preservingly) all the lines in $\tilde{\mu}$ contained in some neighbourhood of $l \in \mathcal{G}$. This is done as follows: since the boxes form a neighbourhood system for $\mathcal{G}$, and $\tilde{\varphi}$ acts via homeomorphisms on $\mathcal{G}$, there are two boxes $B \subset B'$ containing l such that $\tilde{\varphi}(B) \subset B'$. Now $B \cap \tilde{\mu} \subset B' \cap \tilde{\mu}$ consists of some parallel lines, which look like $J \times \mathbb{R} \subset J' \times \mathbb{R}$ for some measured ordered set J' and some subsegment $J \subset J'$: the map φ sends J to J' preserving both the ordering and the measure, and fixing the point corresponding to l: hence it is the identity.

On the other hand, every complementary region of $\tilde{\mu}$ is a finite ideal polygon: hence if $\tilde{\varphi}$ fixes (orientation-preservingly) one side of the polygon, it fixes all the others.

The two properties just listed together easily imply that, since l is fixed, the whole of $\tilde{\mu}$ is. Since the lines are fixed orientation-preservingly, their endpoints are fixed: the map $\tilde{\varphi}$ fixes all the endpoints of all lines in $\tilde{\mu}$, and since these endpoints form a dense subset of $\partial \mathbb{H}^2$ the map $\tilde{\varphi}$ fixes $\partial \mathbb{H}^2$ pointwise and hence φ is trivial in $\mathrm{MCG}(S_g)$ by Proposition 6.5.12. $\square$

The previous proposition shows in particular that pseudo-Anosov elements act freely on $\mathcal{ML} \setminus \{0\}$. We now investigate more closely these mysterious mapping classes.

8.4.6. Pseudo-Anosov elements. A pseudo-Anosov element φ is by definition neither finite order nor reducible. We have just seen that φ acts freely on $\mathcal{ML} \setminus \{0\}$, but this does not prevent it from having some fixed points in $\mathbb{P}\mathcal{ML}$; indeed we now show that there are two fixed points there, one attracting and the other repelling, so that φ looks very much like a hyperbolic isometry on the hyperbolic space.

Theorem 8.4.4. *Let $\varphi \in \mathrm{MCG}(S_g)$ be a pseudo-Anosov element. There are two measured geodesic laminations $\mu_s, \mu_u \in \mathcal{ML}$ and a real*
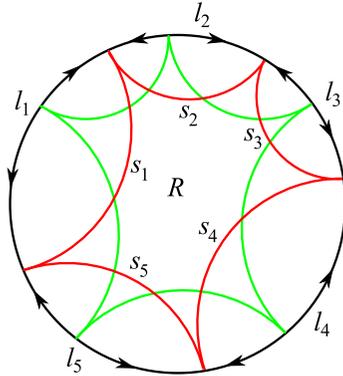
Figure 8.16. The appropriate lift $\tilde{\varphi}$ acts on $\partial\mathbb{H}^2$ with $2k$ fixed points that are alternatively attractive and repelling. By joining the repelling points we find another lamination μ_u fixed by φ. Here $k = 5$.

number $\lambda > 1$ such that

$$\varphi(\mu_s) = \lambda\mu_s, \qquad \varphi(\mu_u) = \frac{1}{\lambda}\mu_u.$$

The laminations μ_s and μ_u are full, and they altogether fill S_g.

Proof. By Brouwer's fixed point theorem, a pseudo-Anosov element φ has a fixed point in $\overline{\mathrm{Teich}(S_g)} \cong D^{6g-6}$ which is (by definition) neither a metric nor a multicurve. Therefore φ fixes a projective measured lamination $[\mu]$ which is full (otherwise φ would be reducible: see the proof of Proposition 8.4.3).

Since $[\mu]$ is a projective class, we have $\varphi(\mu) = \lambda\mu$ for some real number $\lambda > 0$. Up to replacing φ with its inverse φ^{-1} we may suppose that $\lambda \geqslant 1$, and Proposition 8.4.3 shows that $\lambda > 1$. We denote this μ by μ_s.

We now construct μ_u. As in the proof of Proposition 8.4.3, we consider the preimage $\tilde{\mu}_s \subset \mathbb{H}^2$ of μ_s, and after replacing φ with a finite power we may choose a lift $\tilde{\varphi}$ of φ that fixes a complementary polygonal region R of $\tilde{\mu}_s$ and its sides, hence in particular the vertices of R, see Figure 8.16.

The k vertices of R divide $\partial\mathbb{H}^2$ into $\tilde{\varphi}$-invariant arcs $l_1, \ldots, l_k$, corresponding to the sides $s_1, \ldots, s_k$ of R, see Figure 8.16. Since μ_s is full, each s_i is the limit of a sequence of leaves in $\tilde{\mu}_s$ with both endpoints in l_i converging to the endpoints of s_i but distinct from them by Proposition 8.3.10. Since $\lambda > 1$, the map $\tilde{\varphi}$ pushes these leaves towards s_i, so in particular the vertices of R are local attractors for the action of $\tilde{\varphi}$ on $\partial\mathbb{H}^2$, see Figure 8.16.

Since the endpoints p and q of l_i are attractors, the map $\tilde{\varphi}$ fixes at least one point in the interior of l_i, and we show that it cannot fix two: if $\tilde{\varphi}$ fixes r and s in the interior of l_i, then it fixes the box $[p, r] \times [s, q]$ that has non-zero measure, a contradiction since $\lambda > 1$. There is a single fixed point in l_i, and it must be repulsive.

The dynamics of $\tilde{\varphi}$ on $\partial \mathbb{H}^2$ is described in Figure 8.16. The closure of the projection of the k green lines is another invariant geodesic lamination μ_u, which must be full because φ is pseudo-Anosov.

We now prove that μ_u admits some (non unique) transverse measure and can hence be considered as an element of $\mathscr{ML}$. The dynamics shows that for some curve $\gamma \in \mathscr{S}$ the supports of $\varphi^{-k}(\gamma)$ tend to that of μ_u. The sequence $\varphi^{-k}([\gamma]) \in \mathbb{P}\mathscr{ML}$ hence converges on a subsequence to a projective measured geodesic lamination with support μ_u.

In principle, the measure supported by μ_u needs not to be unique, not even up to rescaling: hence the element $[\mu_u]$ may not be uniquely determined and we cannot conclude that it is fixed by φ, unfortunately. However, distinct measures on the same support form obviously a convex cone in the current space $\mathscr{C}$ and hence a closed disc in $\mathbb{P}\mathscr{C}$, the class φ acts on this disc and therefore has a fixed point there by Brouwer's fixed point theorem again.

We have $\varphi(\mu_u) = \lambda' \mu_u$ for some λ'. It is clear that μ_u and μ_s fill S_g altogether, and in particular $i(\mu_u, \mu_s) > 0$. Therefore

$$0 < i(\mu_u, \mu_s) = i(\varphi(\mu_u), \varphi(\mu_s)) = \lambda \lambda' i(\mu_u, \mu_s)$$

gives $\lambda' = \frac{1}{\lambda}$. The proof is complete. $\square$

The laminations μ_s and μ_u are the *stable* and the *unstable* measured geodesic laminations fixed by φ, and λ is the *dilatation* of φ. We now prove a converse to Theorem 8.4.4.

Proposition 8.4.5. *If $\varphi \in \mathrm{MCG}(S_g)$ is such that $\varphi(\mu) = \lambda \mu$ for some full $\mu \in \mathscr{ML}$ and $\lambda > 1$, then φ is pseudo-Anosov.*

Proof. We need to prove that φ is neither finite order nor reducible, that is that φ fixes no non-trivial current $\alpha \in \mathrm{Teich}(S_g) \cup \mathscr{M}$. We have

$$i\big(\varphi^k(\alpha), \mu\big) = i\big(\alpha, \varphi^{-k}(\mu)\big) = \lambda^{-k} i\big(\alpha, \mu\big) \to 0$$

as $k \to +\infty$. Therefore $\varphi(\alpha) \neq \alpha$ unless $i(\alpha, \mu) = 0$, which is excluded since μ is full. $\square$

8.4.7. Examples. We now construct plenty of pseudo-Anosov diffeomorphisms. We need a bit of simple linear algebra.

We say that a matrix or vector is *positive* if all its entries are. A square matrix M with non-negative integral entries is *Perron-Frobenius* if M^k is positive for some $k > 0$.

Proposition 8.4.6. *Every Perron-Frobenius integral matrix M has a positive eigenvector v with some eigenvalue $\lambda > 1$.*
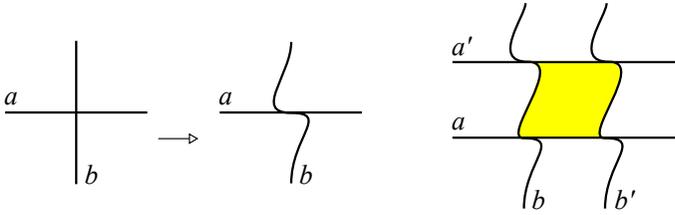
Figure 8.17. We smoothen $\alpha \cup \beta$ to a bigon track that contains both α and β (left). The bigon track may have some bigons as complementary regions (right).

Proof. The matrix M has non-negative entries and hence preserves the standard simplex $\Delta = \{x_i \geqslant 0, \sum x_i = 1\}$, so by Brouwer's fixed point theorem it has a fixed point $v \in \Delta$ there. Since $M^k > 0$ and v is an eigenvector for M^k, we get $v > 0$. Since M^k is integral we get $\lambda^k > 1$ and hence $\lambda > 1$. □

The following examples were constructed by Penner in 1988. Recall that a multicurve α in S_g is *essential* if it contains no parallel components: this holds for instance if α is a simple closed curve or a pants decomposition.

Theorem 8.4.7. *Let α and β be two essential multicurves that altogether fill S_g. Let $\varphi \in \mathrm{MCG}(S_g)$ be any composition of Dehn twists T_a^{+1} and T_b^{-1} where a and b vary among the curves in α and β. If every component a, b of α, β occurs at least once, then φ is pseudo-Anosov.*

The Dehn twists may occur in any order, for instance $\varphi = T_a^2 T_b^{-3} T_{b'}^{-1} T_{a'}$.

Proof. We put α and β in minimal position and smoothen the transverse intersections as in Figure 8.17-(left) to get a *bigon track* τ. A bigon track is like a train track, except that it may contain some complementary bigon as in Figure 8.17-(right). The straightening procedure described in Section 8.3.12 works also in this case (exercise), so every weight system on τ determines a measured geodesic lamination in S_g. (Different weight systems may determine the same geodesic laminations because of the bigons.)

We have $\alpha = a_1 \sqcup \ldots \sqcup a_m$ and $\beta = b_1 \sqcup \ldots \sqcup b_n$. Each a_i and b_j may be represented by assigning weights 1 or 0 to the edges of τ, and these weights form $m+n$ independent (exercise) vectors $v_1, \ldots, v_{m+n}$ in the weights space of τ. Let V be the $(m + n)$-dimensional sub-cone of the weights space generated by $v_1, \ldots, v_{m+n}$ via combinations with non-negative coefficients. Every vector in V models a measured geodesic lamination in S_g.

Let Ω be the $m+n$ square matrix $\Omega_{ij} = i(v_i, v_j)$. We have $\Omega = \left(\begin{smallmatrix} 0 & H \\ K & 0 \end{smallmatrix}\right)$. The crucial point here is that both T_{a_j} and $T_{b_k}^{-1}$ act on V like the matrix

$$Q_i = I + D_i\Omega$$

where I is the identity and D_i has 1 on the i-th entry of the diagonal and 0 everywhere else. The map φ therefore acts on V as a product $M = Q_{i_1} \cdots Q_{i_h}$ of non-negative matrices. Since each $i = 1, \ldots, m+n$ occurs at least once as an index, we may deduce (exercise) that M is Perron-Frobenius. Therefore M has a positive eigenvector v with eigenvalue $\lambda > 1$.

The positive eigenvector v determines an element $\mu \in \mathcal{ML}$ such that $\varphi(\mu) = \lambda\mu$. We leave to the reader the proof that μ is full, and we conclude using Proposition 8.4.5. $\qquad\square$

8.4.8. References. The material contained in this chapter is well known to experts, but it is hard to find a source in the literature that contains everything in a fully self-contained way. The whole theory was presented by Thurston in a very nice and readable paper [**55**] that however contained no proof. The most complete book on the subject is then Fathi – Laudenbach – Poénaru [**19**], and this is also the main source that we have used for writing this chapter. Another important source is Casson – Bleiler [**12**].

We have chosen to describe the whole theory using Bonahon's geodesic currents, that originated in the papers [**6, 7**]. To this purpose we have also consulted McMullen [**39**], Aramayona – Leininger [**3**], and Calegari [**9**]. We also borrowed some arguments on train tracks and measured geodesic laminations from a nice and self-contained paper of Hatcher [**27**]. The proof of Proposition 8.4.1 was taken from [**18**, Theorem 6.1]. Theorem 8.4.7 was proved by Penner in [**44**].

Part 3

Three-manifolds

CHAPTER 9

Topology of three-manifolds

The three-manifolds world is topologically much richer than the surfaces realm, while yet not so crazy as the four-manifolds universe. In dimension two a simple homological invariant (the Euler characteristic) suffices to classify topologically all the closed orientable manifolds: at the complete opposite, the closed four-manifolds cannot be classified in any reasonable sense. The three-manifolds lie in the middle: we do not have yet a complete satisfactory picture, but we understand them a good deal.

The rest of this book is devoted to three-manifolds, more specifically to compact orientable three-manifolds, possibly with boundary (the orientability assumption is not essential, but it helps to simplify many arguments). We split their study into some parts. First, we state some universal topological facts, mostly concerning the way surfaces can be contained in three-manifolds. Then we construct some classes of three-manifolds, focusing mostly on Seifert manifolds. After classifying the Seifert manifolds topologically, we assign a geometry to each: there are eight interesting geometries in dimension three, and we introduce them with some care. Finally, we concentrate on the most interesting and beautiful of the eight: hyperbolic geometry.

In this chapter we start to study the topology of three-manifolds. We begin with some algebraic topology, then we show that the connected sum behaves like products of numbers: every closed three-manifold splits in a unique way in a unique list of prime factors. Finally we introduce the important notion of incompressible surface.

9.1. Algebraic topology

The algebraic topology of compact 3-manifolds is not complicated.

9.1.1. Integral homology. In this section all the homology groups are considered over $\mathbb{Z}$. We first note that, for closed orientable 3-manifolds, the fundamental group determines everything.

Proposition 9.1.1. *The homology $H_*(M)$ of a closed orientable 3-manifold M is determined by $\pi_1(M)$.*

Proof. As in every path connected space, the group $H_1(M)$ is the abelianisation of $\pi_1(M)$ and $H^1(M) = \text{Hom}(H_1(M), \mathbb{Z})$, which is hence

isomorphic to $H_1(M)$ modulo its torsion. By Poincaré duality $H_2(M) = H^1(M)$ and $H^2(M) = H_1(M)$. Finally $H_3(M) = \mathbb{Z}$.   □

9.1.2. Homology spheres. A *homology n-sphere* is a closed orientable *n*-manifold M whose homology is the same as that of S^n. That is, we have $H_0(M) = H_n(M) = \mathbb{Z}$ and $H_i(M)$ vanishes for all $0 < i < n$.

By Poincaré duality, a closed 3-manifold M is a homology sphere if and only if $H_1(M)$ vanishes. Since $H_1(M)$ is the abelianization of $\pi_1(M)$, this happens precisely when $\pi_1(M)$ is a *perfect group*, that is a group with trivial abelianization.

In 1900 Poincaré conjectured that every homology 3-sphere should be homeomorphic to S^3. Four years later, he found himself a counterexample by constructing what is known today as *Poincaré's homology sphere*, a closed 3-manifold with a perfect fundamental group of order 120. He then modified his original conjecture by asking whether every simply connected closed 3-manifold should be homeomorphic to S^3. This fact, widely known as *Poincaré's Conjecture*, was proved only in 2002 by Perelman.

We will construct Poincaré's homology sphere (and many more homology spheres) in Chapter 10, and we will discuss the Poincaré Conjecture in Section 12.9, as a part of Thurston's wider *Geometrisation Conjecture*, also proved by Perelman in 2002.

9.1.3. The boundary. The Euler characteristic $\chi(M)$ of a closed odd-dimensional manifold vanishes, and on manifolds with boundary we have the following.

Proposition 9.1.2. *If M is a compact 3-manifold with boundary, then*

$$\chi(M) = \frac{\chi(\partial M)}{2}.$$

Proof. We only use that M has odd dimension n. If M is closed and orientable, we have $\chi(M) = \sum_{i=0}^{n}(-1)^i b_i$ and the Betti numbers b_i and b_{n-i} are equal by Poincaré duality, hence $\chi(M) = 0$. If M is non-orientable then it has an orientable double-cover N and $\chi(N) = d\chi(M)$ on degree-d covers, hence $\chi(N) = 0$ implies $\chi(M) = 0$. If M has boundary then

$$0 = \chi(DM) = 2\chi(M) - \chi(\partial M)$$

where DM is the *double* of M, constructed by taking two identical copies of M and identifying their boundaries in the obvious way.   □

The manifold M has half the Euler characteristic of ∂M, and is also responsible for half of the first homology group of ∂M. More precisely, we are interested in the boundary map

$$\partial \colon H_2(M, \partial M) \longrightarrow H_1(\partial M)$$

and we want to prove that its image is a particular half-dimensional sub-group. Recall from Proposition 1.7.16 that every class in $H_2(M, \partial M)$ is

represented by an oriented properly embedded surface Σ, and ∂ sends $[\Sigma]$ to $[\partial\Sigma]$.

Let M be oriented. The boundary ∂M may be disconnected and inherits an orientation. Recall that $H_1(\partial M) \cong \mathbb{Z}^{2n}$ for some n and $H_1(\partial M)$ is equipped with a symplectic intersection form ω, see Section 6.1.3. A subgroup $L < H_1(\partial M)$ is *lagrangian* if $\omega|_L \equiv 0$.

Exercise 9.1.3. If L is lagrangian then $\mathrm{rk}\, L \leqslant n$.

When $\mathrm{rk}\, L = n$ we say that L has maximal rank.

Proposition 9.1.4. *Let M be an oriented compact 3-manifold with boundary. The image of the map*

$$\partial\colon H_2(M, \partial M, \mathbb{Z}) \longrightarrow H_1(\partial M, \mathbb{Z})$$

is a lagrangian subgroup of $H_1(\partial M, \mathbb{Z})$ of maximal rank.

Proof. Consider the long exact sequence

$$\ldots \longrightarrow H_2(M, \partial M) \xrightarrow{\partial} H_1(\partial M) \xrightarrow{i_*} H_1(M) \longrightarrow \ldots$$

We have two pairings

$$\omega\colon H_1(\partial M) \times H_1(\partial M) \longrightarrow \mathbb{Z},$$
$$\eta\colon H_2(M, \partial M) \times H_1(M) \longrightarrow \mathbb{Z}.$$

The latter is provided by Lefschetz duality and is non-degenerate after quotienting the torsion subgroups. We have

$$\omega(\partial\alpha, \beta) = \eta(\alpha, i_*\beta)$$

for all $\alpha \in H_2(M, \partial M)$ and $\beta \in H_1(\partial M)$: this equality can be proved easily by representing the elements of $H_2(M, \partial M)$ as surfaces (which we can do thanks to Proposition 1.7.16). Now if $\beta = \partial\alpha'$ we get

$$\omega(\partial\alpha, \partial\alpha') = \eta(\alpha, i_*\partial\alpha') = \eta(\alpha, 0) = 0$$

and hence $\mathrm{Im}\, \partial$ is lagrangian. It has maximal rank since $H_2(M, \partial M) = H^1(M)$ and $H_1(M)$ have the same rank: if $\mathrm{rk}\, \mathrm{Im}\, \partial < \frac{1}{2} b_1(\partial M)$, then we get $\mathrm{rk}\, \mathrm{Im}\, i_* > \frac{1}{2} b_1(\partial M)$ and $\mathrm{rk}\, \ker \partial > b_1(M) - \frac{1}{2} b_1(\partial M)$, a contradiction since $\mathrm{Im}\, i_*$ and $\ker \partial$ are η-orthogonal. $\square$

Corollary 9.1.5. *Let M be an oriented compact 3-manifold. We have*

$$b_1(M) \geqslant \frac{b_1(\partial M)}{2}.$$

Proof. The rank of $H_1(M)$ equals that of $H^1(M) = H_2(M, \partial M)$. $\square$

Corollary 9.1.6. *The boundary of a simply connected compact 3-manifold consists of spheres.*

9.1.4. Non-orientable surfaces. Let M be an orientable 3-manifold and $S \subset M$ be a connected surface. If S is orientable, by Proposition 1.1.12 a tubular neighbourhood for S is diffeomorphic to the product $S \times \mathbb{R}$.

The orientable manifold M may also contain a *non*-orientable surface S: for instance the orientable projective space $\mathbb{RP}^3$ contains the non-orientable projective plane $\mathbb{RP}^2$. In that case a tubular neighbourhood of S is diffeomorphic to the unique orientable interval bundle $S \mathbin{\widetilde{\times}} \mathbb{R}$ over S, see Proposition 1.1.12 again. A compact tubular neighbourhood of S is an interval bundle $S \mathbin{\widetilde{\times}} I$, whose boundary is the orientable double cover of S. For instance, the boundary of $\mathbb{RP}^2 \mathbin{\widetilde{\times}} I$ is a sphere.

A non-orientable properly embedded surface $S \subset M$ does not define a homology class in $\mathbb{Z}$, but it defines one in $\mathbb{Z}_2$, that is $[S] \in H_2(M, \partial M; \mathbb{Z}_2)$. As opposite to orientable surfaces, this class $[S]$ is always non-trivial.

Proposition 9.1.7. *Let M be orientable. Every non-orientable surface S determines a non-trivial class $[S] \in H_2(M, \partial M; \mathbb{Z}_2)$. The manifold M contains at most $\dim H_2(M, \partial M; \mathbb{Z}_2)$ disjoint non-orientable surfaces.*

Proof. A tubular neighbourhood $S \mathbin{\widetilde{\times}} I$ of S contains a simple closed loop α intersecting S transversely in one point. The homology class $[S] \in H_2(M, \partial M; \mathbb{Z}_2) = H^1(M; \mathbb{Z}_2)$ sends α to $1 \in \mathbb{Z}_2$ and is hence non-trivial. If $S = S_1 \sqcup \ldots \sqcup S_k$ are all non-orientable, each S_i has its own α_i and therefore the elements $[S_1], \ldots, [S_k] \in H_2(M, \partial M; \mathbb{Z}_2)$ are linearly independent. $\qquad\square$

Corollary 9.1.8. *A simply-connected three-manifold does not contain any closed non-orientable surface.*

9.2. Prime decomposition

In this section we study spheres and discs in three-manifolds. We prove that every sphere in $\mathbb{R}^3$ bounds a disc, and we call *irreducible* a three-manifold that has this property. Then we study the connected sum of three-manifolds, and we show that it behaves like multiplication of natural numbers: every closed oriented three-manifold decomposes uniquely into some *prime* factors.

9.2.1. Balls. In general dimension n, we have used the term *disc* to denote the closed Euclidean disc D^n, and the term *ball* for its open interior. Speaking about three-dimensional spaces, we will henceforth use a more intuitive terminology and call *disc* and *ball* respectively the closed discs D^2 and D^3, and we will use the symbols D and B for them.

Let M be a connected 3-manifold. Theorem 1.1.14 says that all the closed balls $B \subset \mathrm{int}(M)$ in M are isotopic and hence the removal of the interior of B from M produces a new manifold N with boundary, which does not depend on B. We have $\partial N = \partial M \sqcup \partial B$.

The inverse operation consists of *capping off* a boundary component of N diffeomorphic to S^2 by attaching a ball B to it, via some diffeomorphism. This operation depends only on the boundary component that is capped off.

Proposition 9.2.1. *The manifold M obtained by capping off a spherical boundary component of N does not depend on the diffeomorphism chosen.*

Proof. There are only two diffeomorphisms up to isotopy, see Theorem 6.4.5, and they are related by a reflection of B, so they produce diffeomorphic manifolds. □

In dimension 3 we can therefore freely remove and add balls without affecting much the topology of the manifold. In particular, by removing the interior of a ball from S^3 we get another ball B, and by attaching a ball to B we get S^3 back. We must thank Smale's Theorem 6.4.1 for that: the situation in higher dimensions is more complicated.

9.2.2. Connected sums. Connected sums exist in any dimension n, but when $n \leqslant 3$ they may be redefined in slightly simpler terms:

Definition 9.2.2. The *connected sum* $M = M_1 \# M_2$ of two oriented connected 3-manifolds M_1, M_2 is constructed by removing the interiors of two closed balls from M_1 and M_2, and then gluing the two resulting spheres via any orientation-reversing diffeomorphism.

In dimension $n \leqslant 3$ these diffeomorphisms are all isotopic, see Theorem 6.4.5, hence this is a good definition (in arbitrary dimension we require the diffeomorphism to extend to the removed discs). We usually work with oriented 3-manifolds to have a uniquely defined connected sum, but in some cases this is not necessary. An orientable manifold M is *mirrorable* if it admits an orientation-reversing self-diffeomorphism.

Exercise 9.2.3. If M_1 is mirrorable, the manifold $M = M_1 \# M_2$ does not depend (up to diffeomorphisms) on the orientations of M_1 and M_2.

If both M_1 and M_2 are oriented and not mirrorable, it may happen that $M_1 \# M_2$ and $M_1 \# \overline{M_2}$ are not diffeomorphic (here $\overline{M_2}$ is M_2 with the reverse orientation). There is also a boundary-version of connected sum:

Definition 9.2.4. The *∂-connected sum* $M = M_1 \#_\partial M_2$ of two oriented 3-manifolds with boundary is constructed by gluing two discs $D_1 \subset \partial M_1$ and $D_2 \subset \partial M_2$ via an orientation-reversing diffeomorphism.

The operation depends only on the components of $\partial M_1, \partial M_2$ containing D_1, D_2, because of Theorem 1.1.14. The following holds:

$$M \# S^3 = M, \qquad M \#_\partial B = M, \qquad \partial(M_1 \#_\partial M_2) = (\partial M_1) \# (\partial M_2).$$

In the latter equality we suppose that ∂M_1 and ∂M_2 are connected.

Figure 9.1. Non-degenerate points of index 0, 1, and 2.
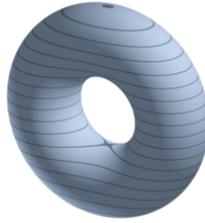

Figure 9.2. The height function on this torus is a Morse function. It has four critical points of index (from bottom to top) 0, 1, 1, 2.

Exercise 9.2.5. Let $M = M_1 \# M_2$ or $M = M_1 \#_\partial M_2$. We have

$$\pi_1(M) = \pi_1(M_1) * \pi_2(M_2).$$

Hint. Use Van Kampen.   □

Corollary 9.2.6. *We have* $H_1(M, \mathbb{Z}) = H_1(M_1, \mathbb{Z}) \oplus H_1(M_2, \mathbb{Z})$.

9.2.3. Irreducible 3-manifolds. We introduce an important definition. Let M be a connected, oriented 3-manifold with (possibly empty) boundary.

Definition 9.2.7. The manifold M is *irreducible* if every sphere $S \subset \text{int}(M)$ bounds a ball.

It is important to recall here that the sphere must be a smooth surface. If we admitted also topological spheres, no 3-manifold would be irreducible: there are "wild" topological spheres inside every ball, as in Figure 1.1.

9.2.4. Alexander theorem. The first 3-manifold to look at is of course $\mathbb{R}^3$. We prove here that $\mathbb{R}^3$ is irreducible.

We need some Morse theory. Let $S \subset \mathbb{R}^3$ be a closed surface and $f(x, y, z) = z$ be the height function. The function f is a *Morse function* for S if $f|_S$ has finitely many critical points, and at each critical point the Hessian of $f|_S$ is non-singular (the Hessian is read on a chart for S, but this definition is chart-independent). The critical point is a local minimum,
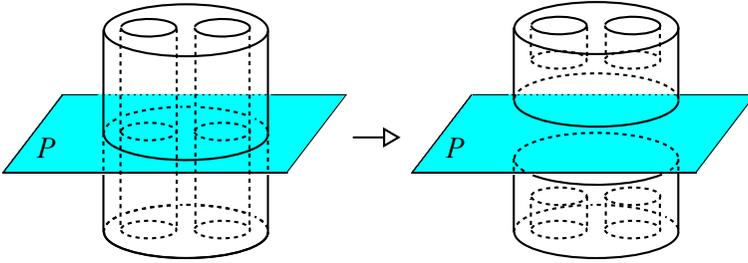
Figure 9.3. The plane P intersects S into circles. Starting from the innermost ones, we cut S along the circles and cap them off by adding pairs of discs. The resulting surface does not intersect P anymore.

a saddle, or a local maximum, according to the signature of the Hessian, see Figure 9.1. These critical points have *index* 0, 1, and 2, respectively.

Exercise 9.2.8. A critical point p for $f|_S$ is non-degenerate if and only if p is a regular (*i.e.* non-critical) point for the Gauss map $S \to S^2$.

Lemma 9.2.9. *Let $S \subset \mathbb{R}^3$ be a closed surface. After rotating S of an arbitrarily small angle, the height function f is a Morse function for S.*

Proof. Consider the Gauss map $\psi \colon S \to S^2$. By Sard lemma there are regular values arbitrarily close to $v = (0, 0, 1)$. Rotate S so that v is a regular value. Now $\psi^{-1}(v)$ is the set of critical points for f and they are all non-degenerate. $\square$

We are now ready to prove the following.

Theorem 9.2.10 (Alexander's Theorem). *The space $\mathbb{R}^3$ is irreducible.*

Proof. Let $S \subset \mathbb{R}^3$ be a 2-sphere. Up to a small rotation we suppose that the height function $f|_S$ is a Morse function, and after a further small rotation we may suppose that the k critical points of $f|_S$ stay at distinct heights $z_1 < \ldots < z_k$. Pick a regular value $u_i \in (z_i, z_{i+1})$ for every $i = 1, \ldots, k-1$. The horizontal plane P at height u_i intersects S transversely into circles. Starting from the innermost ones, we cut S along these circles and cap them off by adding pairs of discs as in Figure 9.3. The resulting surface is disjoint from P.

At every cut a sphere decomposes into two spheres. If we do this for every $i = 1, \ldots, k-1$ we end up with many spheres of the types shown in Figure 9.4, that clearly bound balls in $\mathbb{R}^3$.

Now we reverse the process and undo all the cuts: we prove inductively that at each backward step we have a set of spheres bounding balls (note that the balls are not disjoint!). At each backward step we replace two spheres S_1, S_2 bounding balls B_1, B_2 with one sphere S. Isotope S_1 and
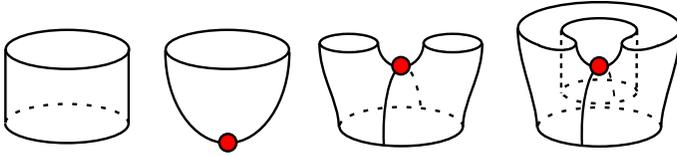
Figure 9.4. After capping off at each u_i we end up with many spheres of these basic types, which clearly bound balls.
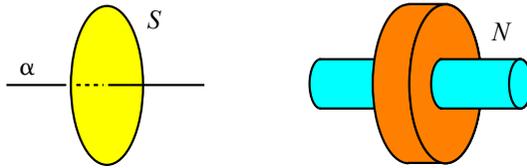

Figure 9.5. If $S \subset M$ is non-separating, there is a simple closed curve α intersecting S transversely in one point: here S is drawn as a disc whose boundary should be collapsed to a point, and α is a line whose endpoints should be identified (left). Pick two tubular neighbourhoods of S and α and consider their union N (right).

S_2, so that they intersect in a disc D. If the interiors of B_1 and B_2 are disjoint, then S bounds the ball $B_1 \cup B_2$. If they are not disjoint, then one is contained in the other, say $B_1 \subset B_2$ and S bounds the ball $B_2 \setminus \text{int}(B_1)$.   □

Corollary 9.2.11. *Every sphere contained in a sub-ball $B \subset M$ of a 3-manifold M bounds a ball $B' \subset B$.*

Corollary 9.2.12. *Every sphere in S^3 bounds a ball on both sides.*

Proof. Choose two points $p, q \notin S$ on opposite sides with respect to S. We have $S^3 \setminus p = S^3 \setminus q = \mathbb{R}^3$ hence S is contained in $\mathbb{R}^3$ in two ways and bounds a ball in each.   □

Alexander's Theorem generalises the smooth Jordan curve Theorem to dimension 3. The situation in higher dimensions is much more problematic: it is still unknown whether every smooth 3-sphere in $\mathbb{R}^4$ bounds a smooth 4-disc (this is usually called the *Schönflies problem*).

9.2.5. Prime manifolds. A connected sum $M_1 \# M_2$ is *trivial* if either M_1 or M_2 is a sphere.

Definition 9.2.13. A connected, oriented 3-manifold M is *prime* if every connected sum $M = M_1 \# M_2$ is trivial.

Being prime is equivalent to be irreducible, with a single exception.

Proposition 9.2.14. *Every oriented 3-manifold $M \neq S^2 \times S^1$ is prime if and only if it is irreducible.*

Proof. The inverse operation of a connected sum $M = M_1 \# M_2$ consists of cutting along a separating sphere $S \subset M$ and then capping off the two resulting manifolds N_1, N_2 with balls. The capped N_i is S^3 if and only if N_i is a ball. Therefore the connected sum is trivial if and only if S bounds a ball on one side. Therefore M is prime if and only if every separating sphere $S \subset M$ bounds a ball.

If M is irreducible, then it is clearly prime. If M is prime and not irreducible, there is a non-separating sphere $S \subset M$. There is a simple closed curve $\alpha \subset M$ intersecting S transversely in one point as in Figure 9.5-(left). Pick two tubular neighbourhoods of S and α as in Figure 9.5-(right): their union is a manifold N with a boundary sphere $\partial N = S'$. The sphere S' is separating and M is prime, hence S' bounds a ball B on the other side and $M = N \cup B$.

We now prove that $M = S^2 \times S^1$. We embed $S \cup \alpha$ naturally in $S^2 \times S^1$ as $S = S^2 \times y$ and $\alpha = x \times S^1$. Decompose $S^2 = D \cup D'$ in two discs and $S^1 = I \cup I'$ in two intervals. The manifold N also embeds as $S^2 \times I \cup D \times S^1$ and its complement $B = D' \times I'$ is a ball. Therefore $M = S^2 \times S^1$. □

9.2.6. Some irreducible manifolds. The Alexander theorem generates many more examples of irreducible 3-manifolds.

Proposition 9.2.15. *Every compact three-dimensional submanifold $M \subset S^3$ with connected boundary is irreducible.*

Proof. Every sphere $S \subset M$ bounds two balls in S^3. Since ∂M is connected, it is contained in one of them, so the other is contained in M. □

We turn to coverings.

Proposition 9.2.16. *Let $p \colon M \to N$ be a covering of 3-manifolds. If M is irreducible then N also is.*

Proof. A sphere $S \subset N$ lifts to many spheres in M, each bounding at least one ball. Pick an innermost such ball B. We prove that $p(B)$ is a ball with boundary S and we are done.

To do this, note that $p|_{\partial B}$ is a diffeomorphism onto S and $p(\text{int}(B))$ is disjoint from S since B is innermost. This implies that $p|_B \colon B \to p(B)$ is a covering and since it has degree one on S it is a diffeomorphism. □

Corollary 9.2.17. *Elliptic, flat, hyperbolic 3-manifolds are irreducible.*

Proof. Their universal covering is diffeomorphic to S^3 or $\mathbb{R}^3$. □

Finally, we consider the exception $S^2 \times S^1$. It is not irreducible since it contains a non-separating sphere, but it is prime.
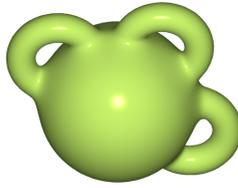
Figure 9.6. A handlebody of genus 3.

Proposition 9.2.18. *The manifold $S^2 \times S^1$ is prime.*

Proof. Let $S \subset S^2 \times S^1$ be a separating sphere: we must prove that it bounds a ball. It separates $S^2 \times S^1$ into two manifolds M and N, and on fundamental groups we get $\mathbb{Z} = \pi_1(M) * \pi_1(N)$. This implies easily that either $\pi_1(M)$ or $\pi_1(N)$ must be trivial: suppose the first.

Since M is simply connected, a copy M' of M lifts to the universal cover $S^2 \times \mathbb{R}$ of $S^2 \times S^1$. We identify $S^2 \times \mathbb{R} = \mathbb{R}^3 \backslash 0$. This copy M' now lies in $\mathbb{R}^3$, and $\partial M' = S^2$ implies that M' is a ball by Alexander theorem. $\square$

9.2.7. Handlebodies and line bundles. We now introduce some basic, but important, compact 3-manifolds with boundary. A *handlebody* is a connected orientable 3-manifold that decomposes into 0- and 1- handles.

Proposition 9.2.19. *The boundary of a handlebody is S_g for some $g \geqslant 0$. Two handlebodies are diffeomorphic if and only if they have the same g.*

Proof. By simplifying handles we may decompose the handlebody into one 0-handle and some g 1-handles, so that the boundary is a genus-g surface, see Figure 9.6. Each 1-handle is attached along a pair of discs; the orientability assumption together with Theorem 1.1.14 easily imply that the result of attaching a 1-handle depends on nothing and the handlebody depends only on g up to diffeomorphisms. $\square$

The *genus* of a handlebody H_g is the genus g of its boundary surface. Proposition 9.2.15 implies the following.

Corollary 9.2.20. *Handlebodies are irreducible.*

Some other simple manifolds are irreducible.

Proposition 9.2.21. *If $g \geqslant 1$ then $S_g \times [0, 1]$ is irreducible.*

Proof. Its universal cover $\mathbb{R}^2 \times [0, 1]$ is irreducible, because its interior is diffeomorphic to $\mathbb{R}^3$. $\square$

Exercise 9.2.22. If $b \geqslant 1$ the manifold $S_{g,b} \times [0, 1]$ is homeomorphic to a handlebody of genus $2g + b - 1 = -\chi(S_{g,b}) + 1$.
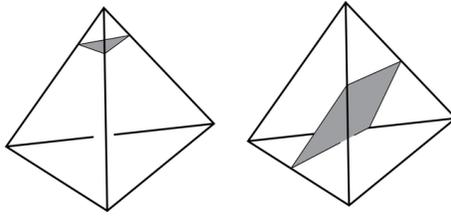
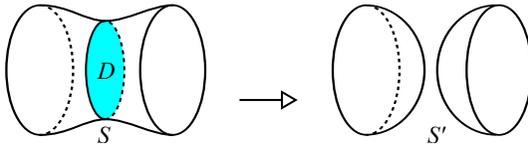Figure 9.7. A normal surface intersects every tetrahedron in triangles or squares.

Figure 9.8. We can *surger* a surface S along a disc D with $\partial D = D \cap S$. The operation consists of removing an annular tubular neighbourhood of ∂D in S and adding two parallel copies of D. We get a new surface S'.

9.2.8. Normal surfaces. Let M be a compact 3-manifold with (possibly empty) boundary. As every smooth compact manifold, M has a triangulation T, made of a certain number t of tetrahedra. We now show that T can be used to treat combinatorially many interesting surfaces in M.

A properly embedded surface $S \subset M$ is *transverse* to T if it is transverse to all its simplexes. In particular S does not intersect the vertices of T, and it intersects every edge, face, and tetrahedron respectively into a finite number of points, curves, and surfaces. Every properly embedded surface $S \subset M$ can be perturbed to be transverse to T.

Definition 9.2.23. A *normal surface* is a properly embedded surface S transverse to T that intersects every tetrahedron into triangles or squares as in Figure 9.7.

Example 9.2.24. For every vertex v of T lying in $\mathrm{int}(M)$ we may take a small sphere centred at v that intersects every incident tetrahedron in a small triangle: we get a normal sphere. If v lies in ∂M we get similarly a normal disc. A surface of this type is called *vertex-linking*. Vertex-linking spheres are not very interesting since they bound balls.

Exercise 9.2.25. A connected normal surface is vertex-linking if and only if it consists of triangles only (no squares).
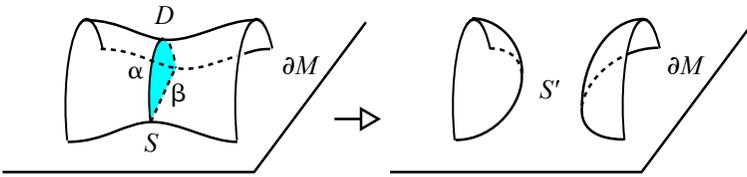
Figure 9.9. We can also surger a surface S along a disc D touching the boundary in a segment. The result is a new properly embedded surface S'.
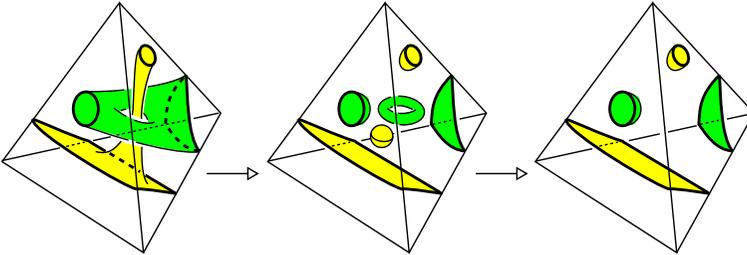

Figure 9.10. Starting from the innermost curves in $\partial\Delta \cap S$ we surger along discs so that $S \cap \Delta$ consists only of discs and closed components, which we then remove.

Let $S \subset M$ be a properly embedded, possibly disconnected, compact surface. An *elementary transformation* on S is one of the following moves:

- the removal of a connected component of S contained in some ball;
- let $D \subset \text{int}(M)$ be a disc with $\partial D = D \cap S$: we *surger* S along D as shown in Figure 9.8;
- let $D \subset M$ be a disc with $D \cap (S \cup \partial M) = \partial D = \alpha \cup \beta$, where α is an arc in S and β an arc in ∂M as in Figure 9.9-(left): we *surger* S as shown in the figure.

Every elementary transformation is local, *i.e.* it takes place in a ball. It transforms S into a new surface S'. We fix a triangulation T for M.

Proposition 9.2.26. *Every properly embedded surface $S \subset M$ becomes normal after finitely many isotopies and elementary transformations.*

Proof. Put S transverse to T. It intersects every tetrahedron Δ into surfaces as in Figure 9.10-(left). The intersection $S \cap \partial\Delta$ consists of closed curves: the innermost ones bound discs in $\partial\Delta$, which may be pushed a bit inside Δ and then used to surger S. We do this with all curves (starting with the innermost ones) to transform S as in Figure 9.10, so that at the end $S \cap \Delta$ consists only of discs and closed surfaces. Then we remove the
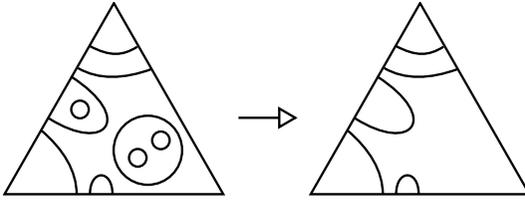
Figure 9.11. The closed curves contained in a triangle are the intersections of small spherical components of S with that triangle. We just remove them.
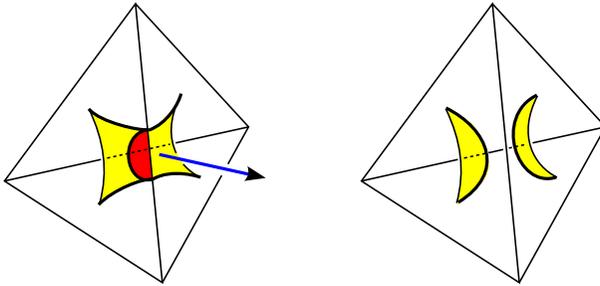

Figure 9.12. If the boundary of a (yellow) disc D intersects an edge e of $\partial\Delta$ twice, we pick two innermost intersections, that must be in opposite directions. There must be a red disc as shown (exercise: use Alexander's theorem), which can be used to push S in the direction indicated by the arrow. The resulting surface is shown in the right. If $e \subset \partial M$, we cannot push S outside ∂M: instead, we surger along the red disc.

closed surfaces: this is an elementary transformation since Δ is a ball. We do this for every tetrahedron Δ.

Now S intersects every tetrahedron Δ in discs. Consider the boundary curve of one such disc: if it is contained in a face as in Figure 9.11-(left), it belongs to one or two discs (depending on whether the face belongs to ∂M or not), which form a component of S contained in a ball that can be removed.

Suppose the boundary curve crosses twice an edge of Δ: if the edge lies in the interior of M, we isotope S as shown in Figure 9.12 to decrease the number of intersections of S with the 1-skeleton of T; if the edge lies in ∂M we simplify analogously by surgering along the red disc shown in the figure. In all cases then we go back to the first step of the algorithm.

After finitely many steps we get a surface S intersecting every tetrahedron Δ in discs, whose boundary curves intersect every edge of Δ at most once. One such disc is either a triangle or a square. $\square$
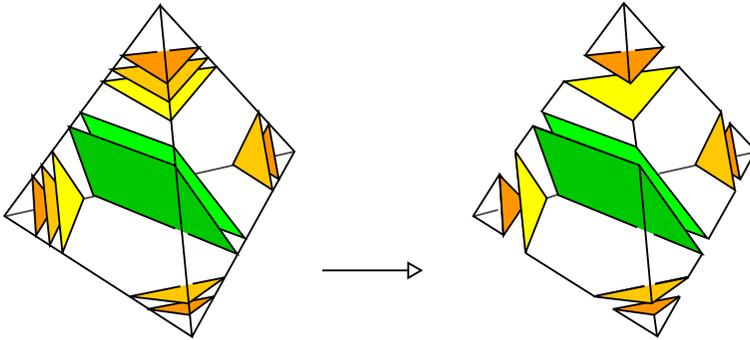
Figure 9.13. If we cut a tetrahedron Δ along triangles and squares we get arbitrarily many prisms with triangular or quadrilateral basis, and at most 6 other pieces (here we get 4 pyramids and two esahedra).

9.2.9. Interesting surfaces cannot have too many components. We know from Proposition 9.1.7 that a compact orientable M cannot contain too many disjoint non-orientable surfaces. It can however contain arbitrarily many orientable surfaces: for instance, small surfaces contained in disjoint balls. A crucial aspect of 3-manifolds theory is that M cannot contain arbitrarily many "interesting" surfaces, as we now see.

Two disjoint connected diffeomorphic surfaces $\Sigma, \Sigma' \subset M$ are *parallel* if they cobound a region diffeomorphic to $\Sigma \times [0,1]$ with $\Sigma = \Sigma \times 0$ and $\Sigma' = \Sigma \times 1$. Two parallel surfaces are obviously isotopic.

Let T be a triangulation of a compact M with (possibly empty) boundary. Let t be the number of tetrahedra in T and set $b_2 = \dim H_2(M, \partial M, \mathbb{Z}_2)$.

Lemma 9.2.27. *Let S be an orientable normal surface. If S has more than $10t + b_2$ components, then two components Σ, Σ' of S are parallel and cobound a $\Sigma \times [0,1]$ which is disjoint from the other components.*

Proof. The complement $M \setminus S$ intersects every tetrahedron Δ into polyhedra: there are many prisms lying between parallel triangles or squares, and at most 6 other pieces, see Figure 9.13. These at most 6 pieces are adjacent to at most $1+1+1+1+3+3 = 10$ triangles and squares.

This implies that, except at most $10t$ of them, the components of S are only adjacent (on both sides) to prisms. These prisms glue to form I-bundles. Therefore at least $b_2 + 1$ components of S are adjacent to I-bundles on both sides. The twisted I-bundles are at most b_2 by Proposition 9.1.7, and each is adjacent to one surface. Therefore at least one surface is adjacent to a product I-bundle $\Sigma \times [0,1]$.   □

We get a topological corollary. A *ball with holes* is a 3-manifold obtained by removing some $k \geqslant 0$ disjoint small open balls from a ball.

A *sphere system* for a 3-manifold M is a surface $S \subset \mathrm{int}(M)$ consisting of disjoint spheres, such that no component of $M \setminus S$ is a ball with holes disjoint from ∂M. (Balls with holes adjacent to spherical components of ∂M are allowed.)

Corollary 9.2.28. *Let M be a compact orientable 3-manifold that does not contain any non-separating sphere. There is a $K > 0$ such that every sphere system in M contains less than K spheres.*

Proof. Pick a triangulation T of M with some t tetrahedra: we prove that $K = 10t + b_2 + 1$ works, with $b_2 = \dim H_2(M, \partial M; \mathbb{Z}_2)$. Suppose by contradiction that there is a sphere system S with $\geqslant K$ spheres. Via isotopies and elementary transformations we transform S into a normal surface S'.

We now examine the effect of elementary transformations in detail. No component of S is contained in a ball, otherwise by Alexander theorem an innermost component of $M \setminus S$ would be a ball. Therefore an elementary transformation cannot cancel a component of S.

A surgery along a disc splits a sphere S_0 into two spheres S_1, S_2. We now prove that by substituting S_0 with either S_1 or S_2 we still get a sphere system. If we push S_1 and S_2 away from S_0 the surfaces S_0, S_1, S_2 altogether bound a ball with two holes B.

By our hypothesis on M the spheres S_0, S_1, and S_2 are separating. Let N_i be the component of $M \setminus (S_0 \cup S_1 \cup S_2)$ adjacent to S_i distinct from B, for $i = 0, 1, 2$. If both N_1, N_2 are balls with holes disjoint from ∂M, then $N_1 \cup N_2 \cup B$ also is, which is excluded. Therefore one, say N_1, is not a ball with holes disjoint from ∂M. On the other side, if $N_2 \cup B \cup N_0$ were a ball with holes disjoint from ∂M, then N_0 would also be (by Alexander's theorem), which is excluded. Therefore by substituting S_0 with S_1 we still get a sphere system.

This proves that the final normal surface S' contains a sphere system with the same number $\geqslant K$ of components as S. Lemma 9.2.27 gives a contradiction.   □

9.2.10. Prime decomposition. We now show that the connected sum operation on oriented three-manifolds behaves like the product of natural numbers: every object decomposes uniquely into prime factors.

Theorem 9.2.29. *Every compact oriented 3-manifold M with (possibly empty) boundary decomposes into prime manifolds:*

$$M = M_1 \# \ldots \# M_k$$

This list of prime factors is unique up to permutations and adding/removing copies of S^3.

Proof. We first show the existence of a decomposition. If M contains a non-separating sphere, then the proof of Proposition 9.2.14 shows
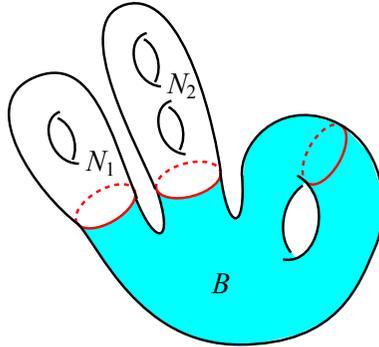
Figure 9.14. A reducing set of spheres (red) for $M = M_1 \# M_2 \# (S^2 \times S^1)$. Here B is a ball with three holes and N_i is M_i with one hole.

that $M = M' \# (S^2 \times S^1)$. Since $H_1(M) = H_1(M') \oplus \mathbb{Z}$, up to factoring finitely many copies of $S^2 \times S^1$ we may suppose that every sphere in M is separating.

If M is prime we are done. If not, it decomposes as $M = M_1 \# M_2$. We keep decomposing each factor until all factors are prime: this process must end, because a decomposition $M = M_1 \# \ldots \# M_k$ gives rise to a system of $k - 1$ spheres, and k cannot be arbitrarily big by Corollary 9.2.28.

We turn to uniqueness. Let

$$M = M_1 \# \ldots \# M_k \#_h (S^2 \times S^1), \qquad M = M'_1 \# \ldots \# M'_{k'} \#_{h'} (S^2 \times S^1)$$

be two prime decompositions with $M_i, M'_j \neq S^2 \times S^1$, so M_i, M'_j are irreducible for all i, j. We say that a set $S \subset M$ of disjoint spheres is a *reducing set of spheres* for the decomposition $M = M_1 \# \ldots \# M_k \#_h (S^2 \times S^1)$ if $M \setminus S$ consists of precisely one M_i with some holes for each i, and some balls with holes disjoint from ∂M. An example is drawn in Figure 9.14. In general, we may construct S by taking the spheres of the prime decomposition, plus one non-separating sphere inside each $S^2 \times S^1$ summand. Similarly, let S' be a reducing set of spheres for the other decomposition.

The first observation we make is that if we add to S any sphere Σ disjoint from S, then we still get a reducing set of spheres for the same decomposition as before. This is because Σ is contained in a holed M_i or S^3, and since M_i is irreducible Σ bounds a ball B there. Therefore by adding Σ we still get the same holed M_i, plus a possibly holed (if $B \cap S \neq \varnothing$) ball B.

We assume S and S' intersect transversely in circles and pick an innermost circle in a component of S bounding a disc $D \subset S$. We surger S' along D, thus substituting a component S'_0 of S' with two spheres $S'_1 \sqcup S'_2$.

We check that the result is another sphere system for the same decomposition. We isotope the spheres S_0', S_1', S_2' so that they are disjoint and cobound a ball with two holes B_2: the system $S' \sqcup S_1' \sqcup S_2'$ is still reducing by the observation above. The removal of S_0' then adds B_2 to the outside of S_0', and this is equivalent to making one more hole there.

After finitely many surgeries we get $S \cap S' = \varnothing$. By the same observation above $S \cup S'$ is a reducing set of spheres for both decompositions: therefore the pieces M_i and M_j' of the decompositions are pairwise diffeomorphic.

Finally we must have $h = h'$ since $M = N \#_h (S^2 \times S^1) = N \#_{h'} (S^2 \times S^1)$ and $H_1(M) = H_1(N) \oplus \mathbb{Z}^h = H_1(N) \oplus \mathbb{Z}^{h'}$. □

This important result is known as the *Prime decomposition Theorem* for 3-manifolds: the existence of a decomposition was proved by Kneser [**33**] in 1929, and its uniqueness by Milnor [**41**] in 1962. In light of this result, topologists have since long restricted their attention to prime 3-manifolds, or almost equivalently to irreducible 3-manifolds.

The strategy of cutting canonically a three-manifold along surfaces has proved successful with spheres, so we now try to do the same with other surfaces. We start by studying properly embedded discs, which of course occur only in manifolds with boundary. It is tempting to guess that discs should behave roughly like spheres, because by doubling a 3-manifold along its boundary we transform properly embedded discs into spheres. We now see that this is indeed the case. As for spheres, we need to distinguish the interesting ones, that we call *essential*, from the others.

9.2.11. Essential discs. Let M be a compact 3-manifold with (possibly empty) boundary. A properly embedded surface $S \subset M$ is *∂-parallel* if it is obtained by slightly pushing inside M the interior of a compact surface $S' \subset \partial M$, possibly with boundary.

In what follows D and S are properly embedded. We say that:

- a sphere $S \subset M$ is *essential* if it does not bound a ball,
- a disc $D \subset M$ is *essential* if it is not ∂-parallel.

Now, the manifold M is:

- *irreducible* if it does not contain essential spheres, and
- *∂-irreducible* if it does not contain essential discs.

To warm up, we show the following. A *solid torus* is a handlebody of genus 1, that is a three-manifold diffeomorphic to $D \times S^1$.

Proposition 9.2.30. *Let M be a compact, irreducible, orientable 3-manifold with boundary, and let $D \subset M$ be an essential disc. Let $\Sigma \subset \partial M$ be the boundary component containing ∂D. Then:*

- *the curve ∂D is non-trivial in Σ;*
- *if Σ is a torus then M is a solid torus.*

Proof. If ∂D bounds a disc $D' \subset \Sigma$ then $D \cup D'$ is a sphere, which bounds a ball because M is irreducible; this ball furnishes a parallelism between D and D' and hence D is ∂-parallel, a contradiction.

If Σ is a torus, by surgerying Σ along D we get a sphere which must bound a ball B. Therefore M is obtained by adding a one-handle to B, that is it is a solid torus.   □

9.2.12. Decomposition along discs. We now show that essential discs behave roughly like essential spheres, in the sense that there is a kind of prime decomposition theorem also for discs. After stating and proving this theorem, we will essentially forget about essential spheres and discs and focus on the 3-manifolds that do not contain them.

Let M be a compact 3-manifold with (possibly empty) boundary. A *disc system* in M is a set of pairwise disjoint non-parallel essential discs.

Proposition 9.2.31. *There is a $K > 0$ such that every disc system in M cannot contain more than K discs.*

Proof. The proof is analogous to that of Corollary 9.2.28   □

We now want to cut irreducible manifolds along essential discs.

Remark 9.2.32. The opposite operation of cutting a manifold along a properly embedded disc is a 1-handle addition.

We now state the analogue of Theorem 9.2.29 in this context. Note that M is irreducible by hypothesis: we have already eliminated the essential spheres, and we now remove the essential discs.

Theorem 9.2.33. *Every compact oriented irreducible 3-manifold M is obtained by adding 1-handles to a finite list*

$$M_1, \ldots, M_k$$

of connected irreducible and ∂-irreducible 3-manifolds. The list is unique up to permutations and adding/removing balls.

Proof. The proof is similar to Theorem 9.2.29. If we cut M along a maximal disc system we get some connected manifolds $M_1, \ldots, M_k$. Each M_i is ∂-irreducible because the disc system is maximal, and it is irreducible because every sphere in $M_i \subset M$ bounds a ball B in M, and B is necessarily contained in M_i.

We prove uniqueness. Pick two sets of discs $S, S' \subset M$ such that by cutting along each we get irreducible and ∂-irreducible components. We prove that these components are the same up to adding/removing balls. We assume that S and S' intersect transversely.

A circle $S \cap S'$ which is innermost with respect to S is bounded by two sub-discs $D \subset S$ and $D' \subset S'$, which together form a sphere: this sphere in turn bounds a ball (because M is irreducible) as in Figure 9.15. We
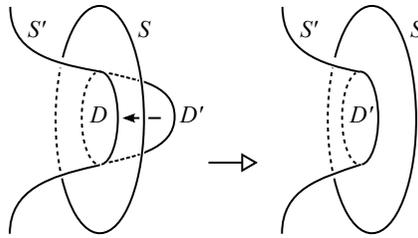
Figure 9.15. The discs D and D' form a sphere which bounds a ball since M is irreducible. We can use this ball to isotope S' away from S.

can use the ball to isotope S' away from S and reduce the intersections in $S \cap S'$. After finitely many isotopies there are no circles in $S \cap S'$.

We are left to consider the arcs in $S \cap S'$. We end precisely as in Theorem 9.2.29. We observe that by adding more discs to S or S' we get a set of discs that ∂-reduces M into the same list of M_i's, plus possibly some balls.

Thanks to an innermost argument, we may surger some disc $D_0 \subset S$ along an arc in $D_0 \cap S'$ and transform it into two discs $D_1 \sqcup D_2$, both disjoint from S and with less intersections with S'. The set $S \sqcup D_1 \sqcup D_2$ is still a reducing set for the M_i's. If we remove D_0 from this new system, in the complement we only remove a ball: hence we can replace D_0 with $D_1 \sqcup D_2$. After finitely many steps we get $S \cap S' = \varnothing$ and therefore $S \cup S'$ is a reducing set producing the same M_i's of S and S', up to balls. $\square$

9.2.13. Projective planes. The previous discussions show how to deal with spheres and discs, and we now turn to the only remaining compact surface with $\chi > 0$: the projective plane. There is not much to say about it.

Proposition 9.2.34. *Every compact irreducible orientable 3-manifold M that contains $\mathbb{RP}^2$ is diffeomorphic to $\mathbb{RP}^3$.*

Proof. The orientable M contains the orientable tubular neighbourhood $N = \mathbb{RP}^2 \mathbin{\widetilde{\times}} I$ of $\mathbb{RP}^2$. The boundary ∂N is the orientable cover S^2 of $\mathbb{RP}^2$. Since M is irreducible, the sphere ∂N bounds a ball B. Therefore M is uniquely determined as the manifold obtained by capping off the boundary sphere of N.

The manifold $\mathbb{RP}^3$ is irreducible (because its cover S^3 is) and contains $\mathbb{RP}^2$, hence we must have $M = \mathbb{RP}^3$. $\square$

This proposition does not hold for non-orientable three-manifolds, which might contain $\mathbb{RP}^2$ with a product neighbourhood $\mathbb{RP}^2 \times I$. These cases are difficult to deal with, so one typically excludes them *a priori* by

requiring the manifold to be $\mathbb{RP}^2$-*irreducible*, which means that it does not contain two-sided projective planes.

9.3. Incompressible surfaces

We have proved that every compact oriented 3-manifold decomposes along essential spheres and discs, into some canonical pieces that do not contain essential spheres or discs anymore.

We would like to pursue this strategy with the next simplest surfaces, namely tori and annuli. To this purpose we define the important notion of *incompressible* surface in a three-manifold, which applies to all surfaces of non-positive Euler characteristic.

9.3.1. Incompressible surfaces. Throughout all this section M denotes a compact orientable 3-manifold with (possibly empty) boundary. Let $S \subset M$ be a properly embedded orientable surface. A *compressing disc* for S is a disc $D \subset M$ with $\partial D = D \cap S$, such that ∂D does not bound a disc in S. With this hypothesis, the surgery in Figure 9.8 is called a *compression*: it transform S into a new surface $S' \subset M$ which is simpler than S.

Proposition 9.3.1. *The surface S' may have one or two components S_i', and $\chi(S_i') > \chi(S)$ for each component.*

Proof. We have $\chi(S') = \chi(S) + 2$. If S' has one component we are done, so suppose $S' = S_1' \sqcup S_2'$. Since ∂D did not bound a disc in S, no S_i' is a sphere, hence $\chi(S_i') \leqslant 1$ that implies $\chi(S_i') > \chi(S)$ for $i = 1, 2$. $\square$

A properly embedded connected orientable compact surface $S \subset M$ with $\chi(S) \leqslant 0$ is *compressible* if it has a compressing disc, and *incompressible* otherwise. See Figure 9.16-(top).

Corollary 9.3.2. *Let $S \subset M$ be any properly embedded orientable surface. After compressing it a finite number of times it transforms into a disjoint union of spheres, discs, and incompressible surfaces.*

Proof. We compress S as much as we can; after finitely many steps we must stop because of Proposition 9.3.1. $\square$

Remark 9.3.3. By definition, an orientable 3-manifold is ∂-irreducible if and only if its boundary consists of incompressible components. (Strictly speaking, a boundary component is not properly embedded and hence cannot be incompressible: we implicitly push it inside the 3-manifold.)

A simple incompressibility criterion is the following.

Proposition 9.3.4. *Let $S \subset M$ be an orientable, connected, properly embedded surface with $\chi(S) \leqslant 0$. If the map $\pi_1(S) \to \pi_1(M)$ induced by inclusion is injective, then S is incompressible.*
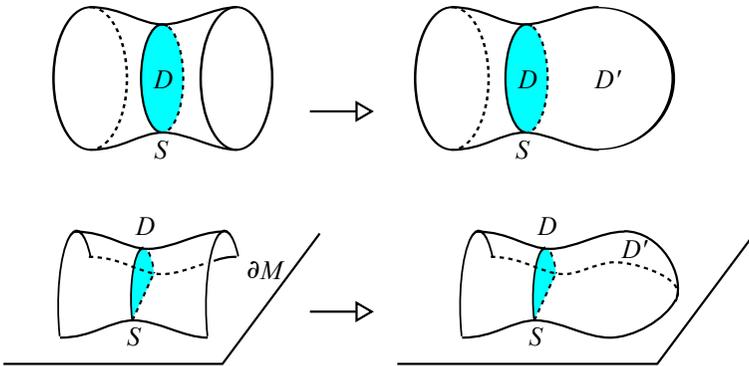
Figure 9.16. A surface S is incompressible (∂-incompressible) if the existence of a disc D as in the top-left (bottom-left) picture implies the existence of another disc $D' \subset S$ as in the top-right (bottom-right) picture. If in addition M is irreducible (∂-irreducible) the two discs D and D' form a sphere (disc) which bounds a ball (is ∂-parallel), and hence by substituting D' with D we get two isotopic surfaces.

Proof. Suppose that a disc D compresses S. Its boundary ∂D is a non-trivial element in $\pi_1(S)$ because it does not bound a disc in S by Proposition 6.3.5, but is clearly trivial in $\pi_1(M)$: a contradiction.   □

The converse is also true, but its proof is much harder! We will complete it at the end of this chapter. For the moment we content ourselves with the following.

Proposition 9.3.5. *If $S \subset M$ is incompressible, every component of ∂S is non-trivial in ∂M.*

Proof. If a component of ∂S is trivial in ∂M, it bounds a disc $D \subset \partial M$ there. By taking an innermost one we get $D \cap S = \partial D$, and by pushing D inside M we find a compressing disc for S.   □

9.3.2. Tori. The first closed surface to look at is the torus.

Proposition 9.3.6. *Let $T \subset M$ be a torus in an irreducible 3-manifold. One of the following holds:*

(1) *T is incompressible,*
(2) *T bounds a solid torus,*
(3) *T is contained in a ball.*

Proof. If T is not incompressible, it compresses along a disc D. The result of the compression is necessarily a sphere $S \subset M$ which bounds a ball B since M is irreducible. If B is disjoint from T, then T bounds a solid
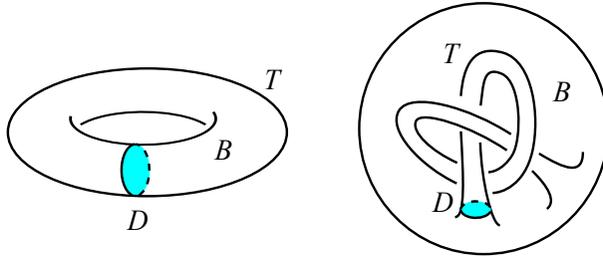
Figure 9.17. A torus in an irreducible 3-manifold is either in-
compressible, or it bounds a solid torus (left), or is contained
in a ball (right).

torus as in Fig.9.17-(left). If B contains T, then case (3) holds as shown
in Figure 9.17-(right).   □

If $M = S^3$ the case (3) can be excluded: if T is contained in a ball B,
then it bounds a solid torus "on the outside" using the complementary ball
$S^3 \setminus B$. We will also exclude the case (1) when we show that S^3 contains
no incompressible surfaces.

9.3.3. ∂-incompressible surfaces. There is of course also a ∂-version
of incompressibility. Let $S \subset M$ be a properly embedded orientable surface
in a 3-manifold M. A ∂-compressing disc for S is a disc D with $\partial D = \alpha \cup \beta$,
where α lies in S and β in ∂M as in Figure 9.9-(left); we also require that
there is no sub-disc $D' \subset S$ with $\partial D' = \alpha \cup \beta'$ and $\beta' \subset \partial S$. The move
in Figure 9.9 is a ∂-compression and transforms S into a surface $S' \subset M$
simpler than S:

Proposition 9.3.7. *The surface S' may have one or two components
S_i', and $\chi(S_i') > \chi(S)$ for each component.*

Proof. We have $\chi(S') = \chi(S) + 1$. If S' has one component we are
done, so suppose $S' = S_1' \sqcup S_2'$. Since α did not bound a disc in S, no S_i'
is a disc, hence $\chi(S_i') \leqslant 0$ that implies $\chi(S_i') > \chi(S)$ for $i = 1, 2$.   □

A properly embedded connected orientable compact $S \subset M$ with
$\chi(S) \leqslant 0$ is ∂-compressible if it has a ∂-compressing disc, and it is ∂-
incompressible otherwise. See Figure 9.16.

Corollary 9.3.8. *Let $S \subset M$ be any properly embedded orientable
surface. After ∂-compressing it a finite number of times it transforms into
a disjoint union of spheres, discs, and ∂-incompressible surfaces.*

9.3.4. Annuli. The first non-closed surface to look at is the annulus.
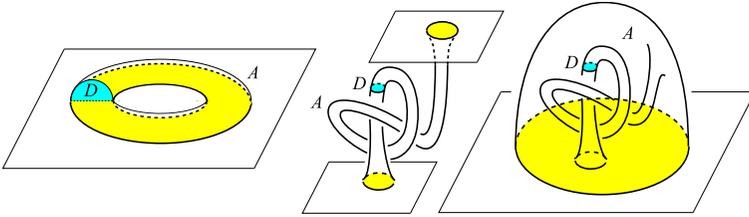Let a *tube* be the tubular neighbourhood of a properly embedded arc.

Figure 9.18. An annulus A in an irreducible and ∂-irreducible 3-manifold is either incompressible and ∂-incompressible, or is parallel to an annulus in ∂M (left), or bounds a tube (centre), or is contained in a ball intersecting ∂M in a disc (right).

Proposition 9.3.9. *Let $A \subset M$ be a properly embedded annulus in an irreducible and ∂-irreducible 3-manifold. One of the following holds:*

(1) *A is incompressible and ∂-incompressible,*
(2) *A bounds a tube,*
(3) *A is parallel to an annulus in ∂M,*
(4) *A is contained in a ball B intersecting ∂M in a disc.*

Proof. If A compresses along a disc D, it transforms into two discs that are parallel to two discs $D_1, D_2 \subset \partial M$ since M is ∂-irreducible. If $D_1 \cap D_2 = \varnothing$ then A bounds a tube as in Figure 9.18-(right); if $D_1 \subset D_2$ then A is contained in a ball B intersecting ∂M in D_2.

If A ∂-compresses along a disc D, it transforms into a disc which is again ∂-parallel and hence A is as in Figure 9.18-(left) or bounds a tube. $\square$

As a corollary, we get a simple criterion for detecting incompressible and ∂-incompressible annuli:

Corollary 9.3.10. *Let $A \subset M$ be a properly embedded annulus in an irreducible and ∂-irreducible 3-manifold. If the components of ∂A are non-trivial and non-parallel in ∂M, the annulus A is incompressible and ∂-incompressible.*

9.3.5. Handlebodies. We study some examples. We investigate the incompressible and ∂-incompressible surfaces in the three-sphere and in the handlebodies. Recall that a (∂)-incompressible surface is always compact, orientable, connected, properly embedded, and with non-positive Euler characteristic by hypothesis.

Proposition 9.3.11. *There are no incompressible surfaces in $\mathbb{R}^3$.*

Proof. Let S be a surface in $\mathbb{R}^3$. By applying as is the proof of Alexander Theorem 9.2.10 we find that S transforms into spheres after surgerying along discs. Therefore S compresses somewhere. $\square$

A high genus closed surface may be embedded in $\mathbb{R}^3$ in a rather complicated way (which may be hard to imagine) and the proposition says that no matter how intricate the surface is, there is always a compressing disc somewhere that one can use to simplify the picture. After finitely many compressions the complicated surface is atomised into some trivial spheres.

Corollary 9.3.12. *There are no incompressible surfaces in S^3.*

Corollary 9.3.13. *There are no incompressible surfaces in the ball B.*

Proof. Use Proposition 9.3.5. □

Corollary 9.3.14. *Every torus in S^3 bounds a solid torus and every properly embedded annulus in B bounds a tube.*

Proof. Apply Propositions 9.3.6 and 9.3.9. Since the manifold is S^3 or B, the cases (3) or (4) easily imply (2). □

The solid torus in S^3 and the tube in B may of course be knotted! We now turn to handlebodies. Recall that a handlebody of genus $g \geq 1$ is irreducible by Corollary 9.2.20, but it is clearly not ∂-irreducible.

Proposition 9.3.15 (Handlebodies). *The genus-g handlebody H_g contains no incompressible and ∂-incompressible surfaces.*

Proof. Suppose that $S \subset H_g$ is incompressible and ∂-incompressible. Pick disjoint essential discs $D_1, \ldots, D_g$ that cut H_g into a ball B. Put S in transverse position with respect to $\sqcup_i D_i$, so that the intersection of S with $\sqcup_i D_i$ consists of circles and properly embedded arcs. Since H_g is irreducible and S is incompressible and ∂-incompressible, all these intersections can be removed by an isotopy of S, as explained below. Then Corollary 9.3.13 gives a contradiction.

The intersections removal goes as follows. Let α be one arc or circle in $S \cap D_j$. Since S is incompressible and ∂-incompressible, there is a disc $D' \subset S$ bounded by α (if α is a circle) or by $\alpha \cup \beta$ with $\beta \subset \partial S$ (if α is an arc) as in Figure 9.16. After substituting α with an innermost intersection of D' with $\sqcup_i D_i$ we may suppose that D' is entirely contained in B and hence we may eliminate α with an isotopy as explained in Figure 9.16. □

We can be more specific on the solid torus H_1.

Proposition 9.3.16. *Every incompressible surface in a solid torus is a ∂-parallel annulus.*

Proof. Every incompressible surface $S \subset H_1$ is ∂-compressible by Proposition 9.3.15. If we ∂-compress it, we get either a disc or an incompressible surface again. Therefore S is constructed iteratively from some discs by a sequence of moves opposite to the ∂-compression, that produce incompressible surfaces at each step. One easily sees that the only incompressible surface that this move can produce at the first step is a ∂-parallel annulus (from a single disc), and then one gets stuck. □
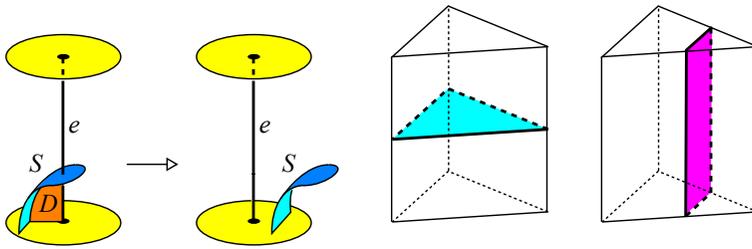
Figure 9.19. A move that decreases the intersection number of S with the vertical edges (left). The surface S intersects every prism into horizontal triangles (centre) or vertical rectangles (right).

Warning 9.3.17. The handlebody H_2 contains many complicated properly embedded incompressible surfaces! However, these are ∂-compressible. The main difference between H_1 and H_2 is that $\pi_1(H_2) = \mathbb{Z} * \mathbb{Z}$ is a free group of rank two and contains many free groups of arbitrarily high rank, so there is a lot of space for π_1-injective incompressible surfaces in H_2 with boundary (whose fundamental group is free).

9.3.6. Line bundles. We now turn to product line bundles. Recall that S_g is the closed orientable surface of genus g.

Proposition 9.3.18 (Line bundles). *Fix $g \geqslant 1$. The product $M = S_g \times [-1, 1]$ is irreducible and ∂-irreducible. The incompressible and ∂-incompressible surfaces in M are precisely the following (up to isotopy):*

- *the* horizontal *surface $S_g \times 0$,*
- *a* vertical *annulus $\gamma \times [-1, 1]$ for each non-trivial simple closed curve $\gamma \subset S_g$.*

Proof. We know that M is irreducible by Proposition 9.2.21. The horizontal surface and the vertical annuli are both incompressible and ∂-incompressible by Propositions 9.3.4 and 9.3.10. Propositions 9.3.4 also shows that M is ∂-irreducible.

We now prove that every incompressible and ∂-incompressible S is either vertical or horizontal, up to isotopy. We use a version of normal surface theory, with prisms instead of tetrahedra.

A triangulation Δ of S_g determines a decomposition of $S_g \times [-1, 1]$ into prisms. Every vertex v of Δ gives rise to a vertical edge $e = v \times [-1, 1]$. We suppose that S is transverse to the prisms and has minimum intersection number with the vertical edges. The intersection of S with a vertical rectangular face consists of arcs and circles. The simplification argument described below shows that after an isotopy we get only horizontal or vertical arcs (that is, arcs joining opposite sides of the rectangle). By

further isotopies we can also suppose that S intersects every horizontal triangle of the prism into arcs joining distinct sides. Therefore S intersects necessarily every prism into either horizontal triangles or vertical rectangles as in Figure 9.19-(centre) and (right). These pieces glue up to give a horizontal or vertical surface.

Here is the simplification argument. There cannot be arcs with endpoints in two consecutive edges, because they could be removed by an isotopy as in Figure 9.19-(left) contradicting minimality. Analogously there are no arcs with both endpoints in the same vertical edge. Circles and arcs with endpoints in the same horizontal edge are removed as in the proof of Proposition 9.3.15.   □

9.4. Haken manifolds

There are two classes of irreducible closed three-manifolds: those that contain incompressible surfaces, and those that do not. Both classes are very important and contain a wealth of interesting manifolds.

The manifolds belonging to the first class are called *Haken manifolds* and are somehow easier to study, because they can be cut into balls via a standard procedure called *hierarchy*: we cut the manifold along an essential surface, then along another, and we iterate until we get balls. We study these manifolds here. We also prove the converse of Proposition 9.3.4, that is that a closed surface is incompressible if and only if it is π_1-injective, see Theorem 9.4.14.

9.4.1. Introduction. If not otherwise mentioned, all the 3-manifolds M we will consider in this section will be connected, compact, oriented and with (possibly empty) boundary. We introduce a definition.

Definition 9.4.1. A *Haken manifold* is a compact, connected, oriented 3-manifold M with (possibly empty) boundary, which is irreducible, ∂-irreducible, and contains an incompressible and ∂-incompressible surface.

The reader should not be frightened by the abundance of adjectives: this definition is really clever because it summarises various reasonable hypothesis in a unique word. The rest of this chapter is mainly devoted to the study of Haken manifolds. We start by looking at their boundaries.

Proposition 9.4.2. *Every boundary component X of a Haken manifold M has $\chi(X) \leqslant 0$ and is incompressible.*

Proof. No component X of ∂M is a sphere: if it were so, it would bound a ball B and we would have $M = B$, contradicting Corollary 9.3.13. Hence $\chi(X) \leqslant 0$ and X is incompressible because M is ∂-irreducible.   □

We now prove that there are plenty of Haken manifolds. We start with a general proposition.

Proposition 9.4.3. *Let M be an oriented, compact, irreducible, and ∂-irreducible 3-manifold with (possibly empty) boundary. Every non-trivial homology class $\alpha \in H_2(M, \partial M; \mathbb{Z})$ is represented by a disjoint union of incompressible and ∂-incompressible oriented surfaces.*

Proof. Every class α is represented by a properly embedded oriented surface S by Proposition 1.7.16. A compression as in Figure 9.8 and 9.9 does not alter the homology class of the surface: indeed in homology we have $S' - S = \partial B$ where $B = D \times [-1, 1]$ is a tubular neighbourhood of the compressing disc D. Hence $[S'] = [S] = \alpha$.

We compress S until its connected components are either incompressible and ∂-incompressible surfaces, discs, or spheres. Since M is irreducible and ∂-irreducible, discs and spheres bound balls and are hence homologically trivial, so they can be removed.   □

Corollary 9.4.4. *Let M be oriented, compact, irreducible, and ∂-irreducible. If $H_2(M, \partial M; \mathbb{Z}) \neq \{e\}$ then M is Haken.*

Corollary 9.4.5. *Let M be oriented, compact, irreducible, and ∂-irreducible. If $\partial M \neq \varnothing$ and $M \neq B$, then M is Haken.*

Proof. If ∂M contains a sphere, it bounds a ball B and hence $M = B$. Otherwise $H_1(\partial M)$ has positive rank, and hence $H_2(M, \partial M) = H^1(M)$ also has positive rank by Corollary 9.1.5.   □

We recall that every compact orientable 3-manifold decomposes along spheres and disc into irreducible and ∂-irreducible pieces. If one such piece has non-empty boundary then either it is ball, or it is Haken.

The following lemma will be useful soon. It says that every Haken manifold contains an interesting "spanning" surface S, that touches all the boundary components (the spanning surface need not to be connected).

Lemma 9.4.6. *Every Haken manifold M contains an oriented surface S, whose components are incompressible and ∂-incompressible, such that $[\partial S \cap X] \in H_1(X, \mathbb{Z})$ is non-trivial for every boundary component X of M.*

Proof. We have $\partial M = X_1 \sqcup \ldots \sqcup X_k$ with $\chi(X_i) \leqslant 0$ for all i. Proposition 9.1.4 says that the image of

$$\partial \colon H_2(M, \partial M, \mathbb{Z}) \longrightarrow H_1(\partial M, \mathbb{Z})$$

is a lagrangian subgroup L of maximal rank. We have

$$H_1(\partial M, \mathbb{Z}) = \oplus_{i=1}^{k} H_1(X_i, \mathbb{Z}).$$

The decomposition is orthogonal with respect to the symplectic intersection form ω. There is an $\alpha \in L$ whose projection to each $H_1(X_i, \mathbb{Z})$ is non-trivial: if not, the space L would be ω-orthogonal to some $H_1(X_i, \mathbb{Z})$ and hence contained in a smaller symplectic subspace, a contradiction since L has maximal rank. Pick any incompressible and ∂-incompressible surface S such that $\partial[S] = \alpha$.   □

9.4.2. Normal surfaces. On Haken manifolds, incompressible surfaces are efficiently detected by normal surfaces.

Proposition 9.4.7. *Let M be Haken and T be a triangulation for M. Every compact surface $S \subset M$ whose components are all incompressible and ∂-incompressible is isotopic to a normal surface.*

Proof. We know that S becomes normal after surgeries as in Figure 9.8 or Figure 9.9. These surgeries are actually isotopies since S is incompressible and ∂-incompressible, and M is irreducible and ∂-irreducible (see Figure 9.16). □

Note that in the proof of Proposition 9.4.7, as in many other proofs, it is crucial that M be irreducible and ∂-irreducible. We can now apply Lemma 9.2.27 to get the following.

Corollary 9.4.8. *Let M be a Haken manifold. There is a $K > 0$ such that every set S of pairwise disjoint and non-parallel incompressible and ∂-incompressible surfaces in M consists of at most K elements.*

Our aim is now to cut a Haken manifold iteratively along incompressible and ∂-incompressible surfaces. The two-dimensional analogue to keep in mind is the following: every surface S_g of genus $\geqslant 2$ can be cut into pairs-of-pants; a pair-of-pants is a quite simple surface, but we are still not satisfied and we further cut it along three arcs into two discs (two hexagons). We have constructed a two-step decomposition of S_g into discs: this is what we would like to extend from two to three dimensions.

9.4.3. Cutting along surfaces. When we cut a 3-manifold along an incompressible surface, some nice properties of the manifold are preserved.

Proposition 9.4.9. *Let M be compact and irreducible, and $S \subset M$ be either an essential disc or an incompressible surface. Let M' be obtained by cutting M along S. The following holds:*

- *the manifold M' is irreducible;*
- *a closed $\Sigma \subset M'$ is incompressible in $M' \iff$ it is so in M.*

Proof. We prove that M' is irreducible. Let $\Sigma \subset M'$ be a sphere. Since M is irreducible, the sphere Σ bounds a ball $B \subset M$. The ball B cannot contain S because all surfaces in a ball are compressible. Therefore $B \subset M'$ and M' is irreducible.

To prove the second assertion, we show that Σ has a compressing disc D in M if and only if it has one in M'. If D lies in M' then of course it lies also in M. Conversely, suppose D lies in M. Put D in transverse position with respect to S and pick an innermost intersection circle in D, bounding a disc $D' \subset D$. Since D' cannot compress S, and since M is irreducible, the disc D' can be isotoped away from S. This simplifies $D \cap S$ and after finitely many steps we get $D \cap S = \varnothing$ and hence $D \subset M'$. □

Corollary 9.4.10. *If we cut a Haken 3-manifold along a closed incompressible surface, we get a disjoint union of Haken 3-manifolds.*

The following consequence is also interesting.

Corollary 9.4.11. *Let M be compact with non-empty boundary. If M is irreducible, then either it is a handlebody or it contains a closed incompressible surface.*

Proof. The manifold M decomposes along essential discs into manifolds $M_1, \ldots, M_k$ that are irreducible and ∂-irreducible. If every M_i is a ball then M is a handlebody. If M_i is not a ball, the closed surface $S = \partial M_i$ is not a sphere and is incompressible in M_i, and hence also in M by Proposition 9.4.9. □

9.4.4. Hierarchies. We want to use incompressible surfaces to cut every Haken manifold into simpler pieces. The procedure goes as follows.

A *hierarchy* for a Haken 3-manifold M is a sequence of 3-manifolds

$$M = M_0 \overset{S_0}{\rightsquigarrow} M_1 \overset{S_1}{\rightsquigarrow} M_2 \overset{S_2}{\rightsquigarrow} \ldots \overset{S_{h-1}}{\rightsquigarrow} M_h$$

where each M_{i+1} is obtained cutting M_i along a properly embedded (possibly disconnected) surface $S_i \subset M_i$, such that the following holds:

- every component of S_i is an incompressible and ∂-incompressible surface or an essential disc, for all i;
- the final manifold M_h consists of balls.

The number h is the *height* of the hierarchy. We now show that every Haken manifold can be "atomised" into balls in three steps.

Theorem 9.4.12. *Every Haken manifold has a hierarchy of height* 3.

Proof. Let S_0 be a maximal family of pairwise disjoint and non-parallel closed incompressible surfaces in M, which exists by Corollary 9.4.8. We cut $M_0 = M$ along S_0 and get M_1.

Every connected component M_1^i of M_1 is Haken by Corollary 9.4.10. By Lemma 9.4.6 for every i there is a "spanning" surface $S_1^i \subset M_1^i$ made of incompressible and ∂-incompressible components that intersects every boundary component of M_1^i. We cut M_1 along the spanning $S_1 = \sqcup S_1^i$ and get a new manifold M_2.

We now prove that M_2 contains no closed incompressible surface. Indeed, if $\Sigma \subset M_2$ were closed and incompressible, then it would be so also in M by Proposition 9.4.9. Since S_0 is maximal, the surface Σ would be parallel to a component of S_0, that is it would cobound a $\Sigma \times [0, 1]$ with it. Since the spanning surface S_1 intersects all the boundary components of M_1, a component of S_1 would be contained in $\Sigma \times [0, 1]$ and would intersect the boundary only on the side of S_0: this is excluded by Proposition 9.3.18 (products do not contain incompressible and ∂-incompressible surfaces with boundary only on one side).
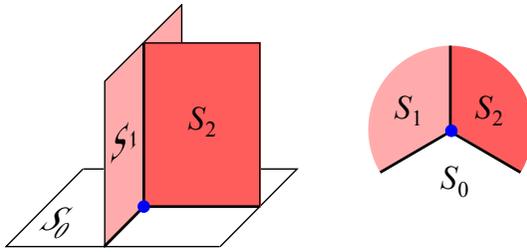
Figure 9.20. The strata S_0, S_1, and S_2 intersect in vertices (left). The boundary of each ball $B \subset M_2$ is tessellated into domains belonging to S_0, S_1, or S_2. Three of them intersect at a 3-valent vertex (right).

Every component of M_2 is a handlebody by Corollary 9.4.11. We cut it along a set S_2 of essential discs to get balls. $\qquad\square$

Hierarchies may be used to prove theorems on Haken manifolds. We will now use them to prove that incompressible surfaces must be π_1-injective: we need a preliminary discussion and a lemma.

In our hierarchy of height 3, the surfaces in S_0 are closed, the spanning surfaces in S_1 have boundary, and S_2 consists of discs. To simplify notations, we redefine S_0 as $\partial M \cup S_0$.

It is convenient to consider all the surfaces S_0, S_1, and S_2 inside M, without cutting M along them. With this perspective ∂S_1 is glued to S_0 and ∂S_2 is glued to $S_0 \cup S_1$, via transverse maps. Every intersection $S_0 \cap S_1 \cap S_2$ is a *vertex* as in Figure 9.20-(left). The space $X = S_0 \cup S_1 \cup S_2$ is a two-dimensional cell complex whose complement in M is a union of open balls: such an object X is usually called a *spine* for M.

We say that our hierarchy of height 3 is *minimal* if the essential discs in S_2 are chosen to minimise the total number of vertices. We can of course suppose that the hierarchy is minimal.

The final manifold M_3 is the abstract closure of $M \setminus (S_0 \cup S_1 \cup S_2)$ and consists of balls. The boundary of every such ball B is tessellated into domains belonging to S_0, S_1, or S_2. Three domains intersect at 3-valent vertices as in Figure 9.20-(right). We say that B is *essential* if every simple closed curve $\gamma \subset \partial B$ transverse to the tessellation and intersecting $\leqslant 3$ domains is the boundary of a disc $D \subset \partial B$ intersecting only these domains: see Figure 9.21.

Lemma 9.4.13. *If the hierarchy is minimal, every ball in M_3 is essential.*

Proof. Let $\gamma \subset \partial B$ intersect $k \leqslant 3$ domains. The curve γ obviously bounds a properly embedded disc $D' \subset B$. If $k = 1$, the curve is entirely contained in S_i for some i. Since S_0 and S_1 are incompressible and S_2 consists of discs, the curve γ bounds a disc D also in S_i. Since S_1 is
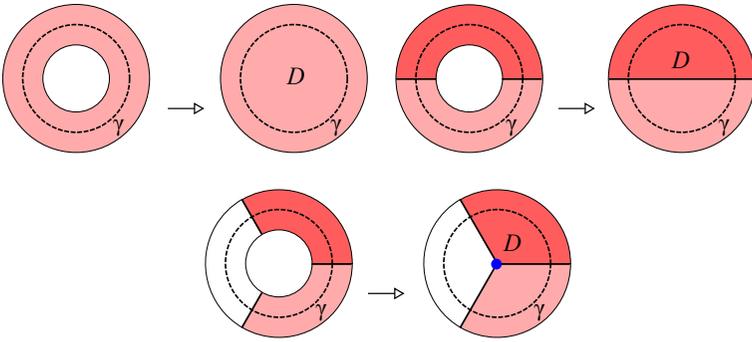
Figure 9.21. If γ intersects at most three domains, it bounds a disc D intersecting only these domains, as shown here.
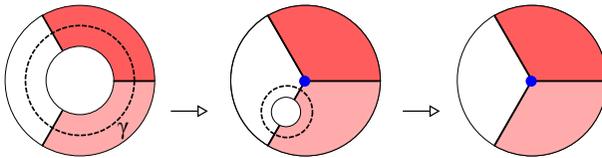

Figure 9.22. By minimality the region in S_2 is incident to a single vertex and we can slide γ through it. Then we resume to the $k = 2$ case.

incompressible and S_2 consists of essential discs, the boundary components of S_j with $j > i$ are not attached in the interior of D, so D is entirely contained in the domain containing γ and we are done.

If $k = 2$, the curve γ is contained in $S_i \cup S_j$ for some $i < j$. If $(i, j) = (0, 1)$ we use that S_1 is ∂-incompressible to get a disc D as in Figure 9.21-(top-right). If $j = 2$, then γ cuts a disc in S_2 into two discs $D_1 \cup D_2$.

We have three half-discs D', D_1, D_2 intersecting in an arc, all contained in a handlebody $H \subset M_2$. Recall that $H \setminus S_2$ consists of balls. If we replace $D_1 \cup D_2$ by either $D_1 \cup D'$ or $D_2 \cup D'$ (say, the first) we still get a disc system that cuts H into balls, and hence another hierarchy. By minimality $D_2 = D$ is adjacent to no vertices, i.e. it is as in Figure 9.21-(top-right) and we are done.

The case $k = 3$ is analogous: the curve γ cuts a disc in S_2 into two parts $D_1 \cup D_2$ and we may replace it with $D_1 \cup D'$. By minimality D_2 is incident to at most one vertex v: we isotope γ through v as in Figure 9.22, and then conclude using the $k = 2$ case.   □
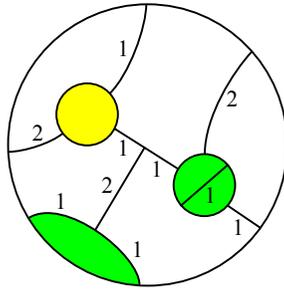
Figure 9.23. The counterimage of $S_0 \cup S_1 \cup S_2$ along f.
An edge with label i goes to S_i (unlabelled edges go to S_0).
Here there is a 0-gon (yellow) and three 2-gons (green). The
correct notion of k-gon should be clear from the picture.

9.4.5. Dehn's Lemma. Proposition 9.3.4 says that an orientable con-
nected properly embedded π_1-injective surface $S \subset M$ with $\chi(S) \leqslant 0$ is
incompressible. We can finally prove the converse, at least for closed sur-
faces.

The following result is often proved as a corollary of a famous topo-
logical fact called the *Dehn Lemma*. We do not state Dehn's Lemma here,
and we prove directly the following using hierarchies.

Theorem 9.4.14. *Let M be a compact oriented 3-manifold. A con-
nected oriented closed $S \subset M$ with $\chi(S) \leqslant 0$ is incompressible if and only
if the induced map $i_* \colon \pi_1(S) \to \pi_1(M)$ is injective.*

Proof. We know one implication from Proposition 9.3.4; here we sup-
pose that S incompressible and prove that $i_* \colon \pi_1(S) \to \pi_1(M)$ is injective.

The decomposition of M into irreducible and ∂-irreducible factors is
made by cutting M along essential spheres and discs transverse to S. Since
S is incompressible, we may surger S along these spheres and discs without
altering i_*, so that S is disjoint from them and hence contained in a single
factor. Therefore we may suppose M is irreducible and ∂-irreducible.

Now M contains the incompressible S and is hence Haken. Theorem
9.4.12 furnishes a hierarchy of height $\leqslant 3$. We may suppose that S_0 is a
maximal system of closed incompressible surfaces containing S and that
the hierarchy is minimal. The balls in M_3 are essential by Lemma 9.4.13.

Suppose by contradiction that i_* is not injective: there is a loop
$\gamma \colon S^1 \to S$ which is homotopically trivial in M but not in S. The triv-
iality in M furnishes a continuous map $f \colon D^2 \to M$ which extends γ. We
homotope f to a smooth map, transverse to all the strata of $S_0 \cup S_1 \cup S_2$.

By transversality, the counterimage $f^{-1}(S_0 \cup S_1 \cup S_2)$ is a graph in
D^2 as in Figure 9.23, which divides D^2 into regions. The graph is itself
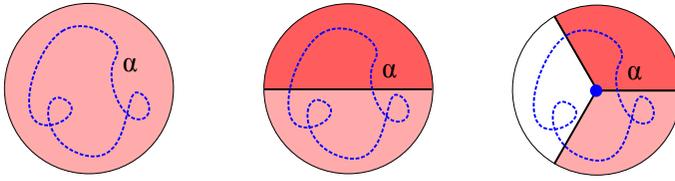
Figure 9.24. The immersed curve α intersects $k \leqslant 3$ regions and is hence of one of these types (because B is essential).

a hierarchy, with edges of type $0, 1, 2$ attached iteratively. An easy Euler characteristic argument shows that at least one region R is a k-gon with $k \leqslant 3$, see Figure 9.23. The region R is mapped inside a ball $B \subset M_3$, and ∂R is mapped to an immersed curve $\alpha \subset \partial B$ intersecting $k \leqslant 3$ domains of the tessellated ∂B.

Since B is essential, the curve α is of one of the types shown in Figure 9.24. In all cases we may slide the disc $f(R)$ away from B and decrease the number of regions in D by destroying R. After finitely many homotopies of this kind we get $f(D) \cap (S_0 \cup S_1 \cup S_2) = \varnothing$ and hence $f(D)$ is entirely contained in a ball $B \subset M_3$. Therefore γ is trivial in S, a contradiction. $\quad\square$

Corollary 9.4.15. *A Haken manifold has infinite fundamental group.*

Proof. It contains the fundamental group of a closed surface with $\chi \leqslant 0$, which is infinite. $\quad\square$

Corollary 9.4.16. *Elliptic 3-manifolds are not Haken.*

We will see in the subsequent chapters that every flat 3-manifold is Haken. Hyperbolic 3-manifolds may or may not be Haken.

9.4.6. Essential surfaces. Topologists sometimes use the term "essential" to summarise various reasonable notions in a single word. We already know what an essential disc or sphere is (see Section 9.2.11) and we now turn to surfaces with non-positive Euler characteristic.

Let M be a compact oriented three-manifold and $S \subset M$ be a properly embedded connected compact surface with $\chi(S) \leqslant 0$. We say that S is *essential* if it is incompressible, ∂-incompressible, and not ∂-parallel.

9.4.7. Simple manifolds. Let M be irreducible and ∂-irreducible. We introduce yet some more definitions. We say that

- M is *atoroidal* if it does not contain essential tori,
- M is *acylindrical* if it does not contain essential annuli.

Finally, the manifold M is *simple* if it is atoroidal and acylindrical. We can summarise this definition as follows:

Definition 9.4.17. A compact oriented 3-manifold M with (possibly empty) boundary is *simple* if it does not contain any essential sphere, disc, torus, or annulus.

Many examples come from elliptic and hyperbolic geometry:

Proposition 9.4.18. *Every closed elliptic or hyperbolic 3-manifold M is simple.*

Proof. We know that M is irreducible. The manifold M does not contain incompressible tori because $\pi_1(M)$ does not contain $\mathbb{Z} \times \mathbb{Z}$: if M is elliptic then $\pi_1(M)$ is finite, if it is hyperbolic we use Corollary 4.2.5.   □

The flat geometry is an exception: the three-torus $S^1 \times S^1 \times S^1$ contains many incompressible tori and is hence not simple. For instance, the two-torus $S^1 \times S^1 \times p$ is incompressible (because it is π_1-injective).

Our next goal will be to decompose every irreducible and ∂-irreducible manifold M along some canonical set of essential tori and annuli into some pieces. These pieces will be either simple, or belong to a particular class: the *Seifert manifolds*. We introduce this class in the next chapter.

9.4.8. References. The main sources that we have used for this chapter are an unfinished book of Hatcher [26] and Fomenko–Matveev [20]. The proof of Theorem 9.4.14 through hierarchies is due to Aitchison and Rubinstein [2].

Seifert manifolds

In the previous chapter we have proved various general theorems on three-manifolds, and it is now time to construct examples. A rich and important source is a family of manifolds built by Seifert in the 1930s, which generalises circle bundles over surfaces by admitting some "singular" fibres. The three-manifolds that admit such kind of fibration are now called *Seifert manifolds*.

In this chapter we introduce and completely classify (up to diffeomorphisms) the Seifert manifolds. In Chapter 12 we will then show how to *geometrise* them, by assigning a nice Riemannian metric to each. We will show, for instance, that all the elliptic and flat three-manifolds are in fact particular kinds of Seifert manifolds.

10.1. Lens spaces

We introduce some of the simplest 3-manifolds, the lens spaces. These manifolds (and many more) are easily described using an important three-dimensional construction, called *Dehn filling*.

10.1.1. Dehn filling. If a 3-manifold M has a spherical boundary component, we can cap it off with a ball. If M has a toric boundary component, there is no canonical way to cap it off: the simplest object that we can attach to it is a solid torus $D \times S^1$, but the resulting manifold depends on the gluing map. This operation is called a *Dehn filling* and we now study it in detail.

Let M be a 3-manifold and $T \subset \partial M$ be a boundary torus component.

Definition 10.1.1. A *Dehn filling* of M along T is the operation of gluing a solid torus $D \times S^1$ to M via a diffeomorphism $\varphi \colon \partial D \times S^1 \to T$.

The closed curve $\partial D \times \{x\}$ is glued to some simple closed curve $\gamma \subset T$, see Figure 10.1. The result of this operation is a new manifold M^{fill}, which has one boundary component less than M.

Lemma 10.1.2. *The manifold M^{fill} depends only on the isotopy class of the unoriented curve γ.*

Proof. Decompose S^1 into two closed segments $S^1 = I \cup J$ with coinciding endpoints. The attaching of $D \times S^1$ may be seen as the attaching
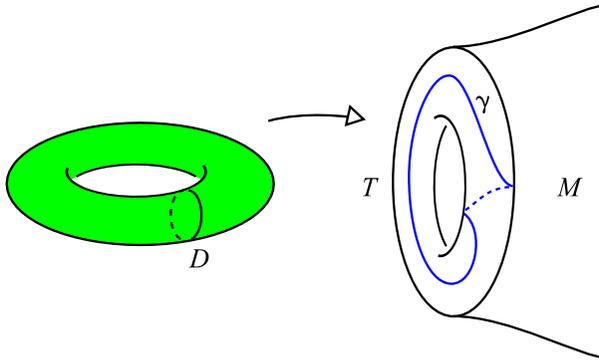
Figure 10.1. The Dehn filling M^{fill} of a 3-manifold M is determined by the unoriented simple closed curve $\gamma \subset T$ to which a meridian ∂D of the solid torus is attached.

of a 2-handle $D \times I$ along $\partial D \times I$, followed by the attaching of a 3-handle $D \times J$ along its full boundary.

If we change γ by an isotopy, the attaching map of the 2-handle changes by an isotopy and hence gives the same manifold. The attaching map of the 3-handle is irrelevant by Proposition 9.2.1.   □

We say that the Dehn filling *kills* the curve γ, since this is what really happens on fundamental groups, as we now see.

The *normaliser* of an element $g \in G$ in a group G is the smallest normal subgroup $N(g) \lhd G$ containing g. The normaliser depends only on the conjugacy class of $g^{\pm 1}$, hence the subgroup $N(\gamma) \lhd \pi_1(M)$ makes sense without fixing a basepoint or an orientation for γ.

Proposition 10.1.3. *We have*

$$\pi_1(M^{\text{fill}}) = \pi_1(M)/_{N(\gamma)}.$$

Proof. The Dehn filling decomposes into the attachment of a 2-handle over γ and of a 3-handle. By Van Kampen, the first operation kills $N(\gamma)$, and the second leaves the fundamental group unaffected.   □

Let a *slope* on a torus T be the isotopy class γ of an unoriented homotopically non-trivial simple closed curve. The set of slopes on T was indicated by $\mathscr{S}$ in Chapter 7. If we fix a basis (m, l) for $H_1(T, \mathbb{Z}) = \pi_1(T)$, every slope may be written as $\gamma = \pm(pm + ql)$ for some coprime pair (p, q). Therefore we get a 1-1 correspondence

$$\mathscr{S} \longleftrightarrow \mathbb{Q} \cup \{\infty\}$$

by sending γ to $\frac{p}{q}$. If T is a boundary component of M, every number $\frac{p}{q}$ determines a Dehn filling of M that kills the corresponding slope γ.

Different values of $\frac{p}{q}$ typically produce non-diffeomorphic manifolds M^{fill}: this is not always true - a notable exception is described in the next section - but it holds in "generic" cases.

10.1.2. Lens spaces. The simplest manifold that can be Dehn-filled is the solid torus $M = D \times S^1$ itself. The oriented *meridian* $m = S^1 \times \{y\}$ and *longitude* $l = \{x\} \times S^1$ form a basis for $H_1(\partial M, \mathbb{Z})$.

Definition 10.1.4. The *lens space* $L(p, q)$ is the result of a Dehn filling of $M = D \times S^1$ that kills the slope $qm + pl$.

A lens space is a three-manifold that decomposes into two solid tori. We have already encountered lens spaces in the more geometric setting of Section 3.4.10, and we will soon prove that the two definitions are coherent. Since $L(p, q) = L(-p, -q)$ we usually suppose $p \geqslant 0$.

Exercise 10.1.5. We have $\pi_1\big(L(p, q)\big) = \mathbb{Z}/_{p\mathbb{Z}}$.

Proposition 10.1.6. *We have $L(0, 1) = S^2 \times S^1$ and $L(1, 0) = S^3$.*

Proof. The lens space $L(0, 1)$ is obtained by killing m, that is by mirroring $D \times S^1$ along its boundary. The lens space $L(1, 0)$ is S^3 because the complement of a standard solid torus in S^3 is another solid torus, with the roles of m and l exchanged (exercise). □

Exercise 10.1.7. Every Dehn filling of one component of the product $T \times [0, 1]$ is diffeomorphic to $D \times S^1$. Therefore by Dehn-filling both components of $T \times [0, 1]$ we get a lens space.

The solid torus $D \times S^1$ has a non-trivial self-diffeomorphism

$$(x, e^{i\theta}) \longmapsto (xe^{i\theta}, e^{i\theta})$$

called a *twist* along the disc $D \times \{y\}$. The solid torus can also be *mirrored* via the map

$$(x, e^{i\theta}) \longmapsto (x, e^{-i\theta}).$$

Exercise 10.1.8. We have $L(p, q) \cong L(p, q')$ if $q' \equiv \pm q^{\pm 1} \pmod{p}$.

Hint. Twist, mirror, exchange the two solid tori giving $L(p, q)$. □

Remark 10.1.9. The meridian m of the solid torus $M = D \times S^1$ may be defined intrinsically as the unique slope in ∂M that is homotopically trivial in M. The longitude l is *not* intrinsically determined: a twist sends l to $m + l$. The solid torus contains infinitely many non-isotopic longitudes, and there is no intrinsic way to choose one of them.

10.1.3. Equivalence of the two definitions. When $p > 0$, we have defined the lens space $L(p, q)$ in two different ways: as the (q, p)-Dehn filling of the solid torus, and as an elliptic manifold in Section 3.4.10. In the latter description we set

$$\omega = e^{\frac{2\pi i}{p}}, \qquad f(z, w) = (\omega z, \omega^q w)$$

and define $L(p, q)$ as $S^3/_\Gamma$ where $\Gamma = \langle f \rangle$ is generated by f. We now show that the two definitions produce the same manifolds.

Proposition 10.1.10. *The manifold $S^3/_{\langle f \rangle}$ is the (q, p)-Dehn filling of the solid torus.*

Proof. The isometry f preserves the central torus

$$T = \left\{ (z, w) \mid |z| = |w| = \tfrac{\sqrt{2}}{2} \right\}$$

that divides S^3 into two solid tori

$$N^1 = \left\{ (z, w) \mid |z| \leqslant \tfrac{\sqrt{2}}{2}, \ |w| = \sqrt{1 - |z|^2} \right\},$$
$$N^2 = \left\{ (z, w) \mid |w| \leqslant \tfrac{\sqrt{2}}{2}, \ |z| = \sqrt{1 - |w|^2} \right\}.$$

Identify T with $S^1 \times S^1 = \mathbb{R}^2/_{\mathbb{Z}^2}$ in the obvious way, so that $H_1(T) = \mathbb{Z} \times \mathbb{Z}$. The meridians of N^1 and N^2 are $(1, 0)$ and $(0, 1)$. The isometry f act on T as a translation of vector $v = \left(\tfrac{1}{p}, \tfrac{q}{p} \right)$. The quotient $T/_{\langle f \rangle}$ is again a torus, with fundamental domain the parallelogram generated by v and $w = (0, 1)$.

The quotients $N^1/_{\langle f \rangle}$, and $N^2/_{\langle f \rangle}$ are again solid tori. Therefore $S^3/_{\langle f \rangle}$ is also a union of two solid tori. Their meridians are the projections of the horizontal and vertical lines in $\mathbb{R}^2$ to $T/_{\langle f \rangle} = \mathbb{R}^2/_{\langle v, w \rangle}$. In the basis (v, w) these meridians are $pv - qw$ and w respectively. Therefore $S^3/_{\langle f \rangle}$ is a $(-q, p)$-Dehn filling on the solid torus, which is diffeomorphic to the (q, p)-Dehn filling by mirroring the solid torus. $\qquad\square$

Corollary 10.1.11. *We have $L(1, 0) = S^3$ and $L(2, 1) = \mathbb{RP}^3$.*

Proof. We have $f = \mathrm{id}$ and $f = -\mathrm{id}$, correspondingly. $\qquad\square$

10.1.4. Classification of lens spaces. Which lens spaces are diffeomorphic? It is not so easy to answer this question, because many lens spaces like $L(5, 1)$ and $L(5, 2)$ have the same homotopy and homology groups, while there is no evident diffeomorphism between them. A complete answer was given by Reidemeister in 1935, who could distinguish lens spaces using a new invariant, now known as the *Reidemeister torsion*. More topological proofs were discovered in th 1980s by Bonahon and Hodgson. We follow here Hatcher [**26**].

Theorem 10.1.12. *The lens spaces $L(p, q)$ and $L(p', q')$ are diffeomorphic $\Longleftrightarrow p = p'$ and $q' \equiv \pm q^{\pm 1} \pmod{p}$.*
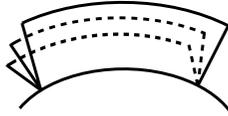
Figure 10.2. We require the p sheets of the spine Δ to lie above every local maximum for $f|_\Sigma$.

Proof. We may suppose $p > 1$. Exercise 10.1.8 furnishes one implication, so we start with a lens space $L(p, q)$ and we prove that there is (up to isotopy) only one torus T dividing $L(p, q)$ into two solid tori: this fact then implies that one can recover p and q intrinsically from $L(p, q)$, up to the stated ambiguity for q (which depends on the chosen longitudes and orientations, and changes by switching the solid tori). So we suppose that there is another torus T'.

Let Σ be the core circle of one solid torus adjacent to T'. If we can isotope Σ inside the torus T, we are done: in that case T' is isotopic to the boundary of a small tubular neighbourhood of Σ, hence both T and T' are cut into two annuli $T = A_1 \cup A_2$ and $T' = A_1' \cup A_2'$ such that $\Sigma \subset A_1$, all four annuli share the same boundaries, and A_2 is contained in the large outside solid torus bounded by T'; the annulus A_2 is incompressible in this solid torus by Proposition 9.3.4 (otherwise Σ would be homotopically trivial, contradicting $p > 1$) and hence ∂-parallel by Proposition 9.3.16, so it is isotopic to either A_1' or A_2', suppose to A_2'; since A_1 is clearly isotopic to A_1' we conclude that T and T' are isotopic.

Our aim is now to prove that Σ can be isotoped inside T. To this purpose we construct two different objects from T' and T, a *spine* Δ and a *foliation* $\mathscr{F}$. The spine $\Delta \subset M$ is built by adding to Σ the meridian D of the other solid torus incident to T', enlarged so that ∂D is glued along Σ like a degree-p covering. Note that $M \setminus \Delta$ is an open ball.

We construct the foliation $\mathscr{F}$ of M by subdividing each solid torus bounded by T into concentric tori, with a central singular circle in each. We represent $\mathscr{F}$ as the level sets of a map $f: L(p, q) \to [0, 1]$ where the singular circles are the extreme levels $f^{-1}(0)$ and $f^{-1}(1)$.

We now put Δ in some good position with respect to $\mathscr{F}$. We first perturb Σ so that it is disjoint from the singular circles of $\mathscr{F}$ and $f|_\Sigma$ is a Morse function with singular points in distinct levels. At every local maximum (minimum) of $f|_\Sigma$ we isotope Δ so that the p sheets of Δ lie above (below) Σ as in Figure 10.2. Finally, we require Δ to be transverse to the two singular circles of $\mathscr{F}$ and $f|_{\Delta \setminus \Sigma}$ to be a Morse function, with critical points at distinct levels (also distinct from those of the critical points of $f|_\Sigma$).
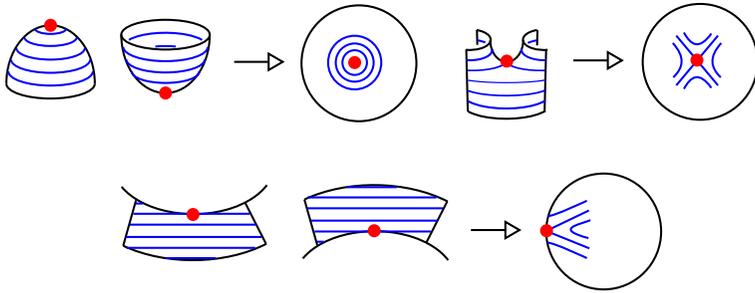
Figure 10.3. The singular points in D can be centres (top left), saddles (top right) and half-saddles (bottom).

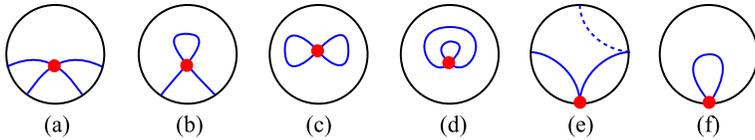(a)   (b)   (c)   (d)   (e)   (f)

Figure 10.4. A singular leaf exiting from a singular point may be of one of these kinds. The dots in (e) indicate that there might be another half-saddle at the other endpoint: this is because there are half-saddles at the same level (precisely p for every local maximum or minimum for $f|_\Sigma$).

The foliation $\mathscr{F}$ induces a singular foliation on Δ that pulls back to D via the map $D \to \Delta$. There are three types of singular points in the foliation, shown in Figure 10.3: centres, saddles, and half-saddles.

A leaf in D incident to a singular point is *singular*. Since we minimised the critical points that may stay at the same level, the singular leaves exiting from a singular point may be only of the six possible kinds shown in Figure 10.4. The saddles and half-saddles of type (a) and (e) are called *essential*, and the others *inessential*. In (a) the singular leaves divide D into four *quarter discs*, and in (e) they cut off two or more *half discs*.

Among all possible good configurations of Δ, we pick one that minimises first the number of critical points in $f|_\Sigma$, and second the number of essential saddles.

Let $D' \subset D$ be a quarter or half-disc not containing any smaller quarter or half-disc. Define $\alpha = D' \cap \partial D$. Suppose first that D' is a half-disc and both endpoints of α are singular. The singular disc D' can contain some singular point of type (b), and it certainly contains one singular point of type (f), see Figure 10.5-(left). (It contains no other singular point except these.) The image of D' in Δ is shown in Figure 10.5-(right). In this case
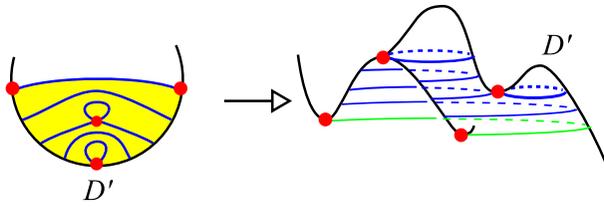
Figure 10.5. The disc D' is a half-disc and α is incident to two singular points.
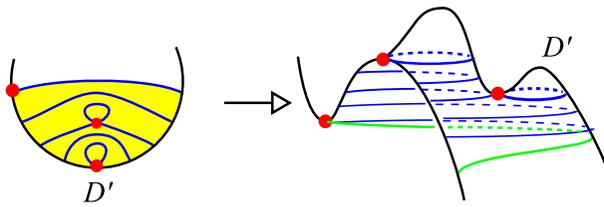

Figure 10.6. The disc D' is a half-disc and α is incident to one singular point.
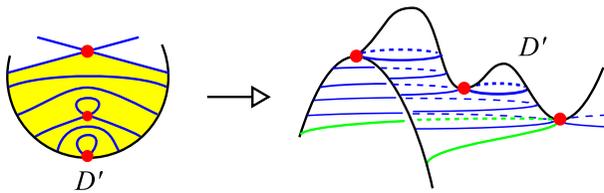
Figure 10.7. The disc D' is a quarter disc.

we can slide α along D' to the green arc shown in the figure which lies entirely in a level torus of f. If we could do this for all the arcs of ∂D cut by singular points, we would happily isotope Σ inside a torus level and hence into T, and we would be done.

There are however two other cases to consider, and both will be excluded by our minimality assumption on Δ. One is that D' may be a half-disc with only one singular endpoint for α, as in Figure 10.6. In that case we can isotope α to the green curve through D', dragging the spine Δ behind. In the new configuration $f|_\Sigma$ has strictly less critical points, a contradiction.

In the last case D' is a quarter disc as in Figure 10.7 and we can isotope α to the green curve dragging Σ as above, but more carefully: we enlarge the other $p - 1$ sheets of Δ by parallel copies of D'. In the new configuration the number of critical points in $f|_{\Sigma}$ is unchanged, but there are strictly fewer essential saddles: at least one is destroyed and no new one is created. (Note that many inessential saddles may be created.) This is also excluded.   □

We have discovered, in particular, that there are closed three-manifolds like $L(5, 1)$ and $L(5, 2)$ that are not diffeomorphic, although their fundamental groups are both isomorphic to $\mathbb{Z}/_{5\mathbb{Z}}$; the two manifolds are both covered by S^3, so they also have isomorphic higher homotopy groups.

10.2. Circle bundles

We now introduce another simple class of 3-manifolds, the orientable circle bundles over some compact surface S. We will discover that there is essentially only one circle bundle if S has boundary, and infinitely many if S is closed, distinguished by an integer called the *Euler number*.

10.2.1. The trivial circle bundle. Let S be a compact connected surface. As every connected manifold, it has a unique orientable line bundle

$$S \times I \quad \text{or} \quad S \mathbin{\tilde{\times}} I$$

depending on whether S is orientable or not. We denote by

$$M = S \times S^1 \quad \text{or} \quad S \mathbin{\tilde{\times}} S^1$$

respectively the double of $S \times I$ and $S \mathbin{\tilde{\times}} I$ along its boundary. If we do not know whether S is orientable or not, we use the symbols $S \mathbin{\overset{(\sim)}{\times}} I$ and $S \mathbin{\overset{(\sim)}{\times}} S^1$ to denote these objects. The manifold $S \mathbin{\overset{(\sim)}{\times}} S^1$ is an orientable circle bundle over S, called the *trivial* one.

10.2.2. Circle bundles with boundary. We start by exploring the case where the base surface S has non-empty boundary: in this case every bundle M over S is a 3-manifold with boundary; the boundary consists of tori, one fibering above each circle in ∂S, because the torus is the unique orientable surface that fibres over S^1.

There is essentially only one bundle over S:

Lemma 10.2.1. *If $\partial S \neq \varnothing$, the orientable circle bundles on S are all isomorphic.*

Proof. Let $N \to S$ be an orientable circle bundle. Decompose S as a disc D with some pairs of disjoint segments (I_i, J_i) in ∂D to be glued. Since D is contractible the restriction of N to D is a product $D \times S^1$ and N is obtained from it by gluing the annuli $I_i \times S^1$ and $J_i \times S^1$ via orientation-reversing fibre-preserving maps. Two such maps are always isotopic (exercise) and hence N is uniquely determined.   □
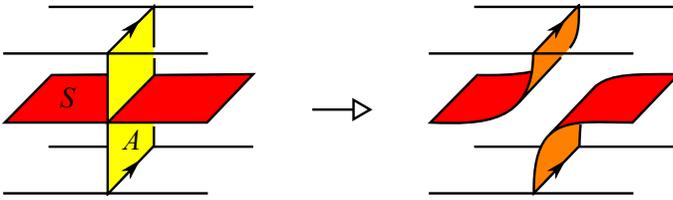
Figure 10.8. A twist along a fibered annulus A modifies a section S as shown.

We now want to study the *sections* of the trivial bundle $M \to S$, because these will be useful in the study of bundles over closed surfaces. We now discover that, although the bundle is trivial, it contains many non isotopic sections, and we want to classify them.

Recall that a *section* of the bundle $\pi\colon M \to S$ is a map $i\colon S \to M$ such that $\pi \circ i = \mathrm{id}$. Since the section i is determined by its image $i(S)$, we simply consider the surface $i(S)$ as a section of π. By construction M is the double of an interval bundle over S and as such it contains the *zero-section* S there. However, this section is not unique in general, not even up to isotopy.

To modify a section, pick a properly embedded arc in S. The arc determines a fibered annulus $A \subset M$ above it, which we may use to *twist* the section as shown in Figure 10.8. This operation modifies the curves $\partial S \subset \partial M$ via two Dehn twists (one positive and one negative) on the tori ∂M along the two curves in ∂A.

By twisting along annuli we may construct all the sections of M:

Lemma 10.2.2. *Two sections of $S \mathbin{\widetilde{\times}} S^1$ are connected by a composition of twists along fibered annuli and fibre-preserving isotopies.*

Proof. Let i_1, i_2 be two sections. Decompose S as a 0-handle D with some rectangular 1-handles attached to ∂D. The bundle is trivial on each handle. Since D is contractible, every map $i\colon D \to S^1$ is homotopic to a constant and we may hence isotope i_1 and i_2 so that they coincide on D.

See a 1-handle as $[-1,1]^2$ with $\{\pm 1\} \times [-1,1]$ glued to D. The sections $i_1, i_2\colon [-1,1]^2 \to S^1$ coincide on $\{\pm 1\} \times [-1,1]$, and up to reparametrising we may suppose that they are constantly 1 there. The section $i_1\colon [-1,1]^2 \to S^1$ defines an element $d \in \pi_1(S^1,1) = \mathbb{Z}$, determined by the loop $t \mapsto i_1(t, u)$, whose homotopy class does not depend on the fixed $u \in [-1,1]$. The section is in fact determined up to isotopy by d, and by twisting along the annulus $0 \times [-1,1]$ we can vary this integer arbitrarily. Therefore after some twists and isotopies the sections i_0 and i_1 are both constantly 1 on 1-handles, and thus coincide everywhere. □

Corollary 10.2.3. *If S has only one boundary component, the boundary of a section of $M = S \mathbin{\widetilde{\times}} S^1$ is a slope in ∂M that does not depend on the section.*

Proof. Distinct sections are connected by finitely many Dehn twists along annuli. One such twist acts on the torus ∂M as a composition of two opposite Dehn twists, which cancel each other. Hence it does not affect the boundary slope of a section. $\square$

Exercise 10.2.4. Prove this corollary using Proposition 9.1.4.

10.2.3. Closed circle bundles. We turn to closed circle bundles. In this section we prove that the oriented circle bundles over a closed surface are parameterised by an integer called the *Euler number*.

We prefer to see the bundles over closed surfaces as Dehn fillings of bundles over surfaces with boundary. Here are the details.

Let S be a compact surface with non-empty boundary. Pick $M = S \mathbin{\widetilde{\times}} S^1$ and fix an orientation for M. Recall that we denote by S the zero-section of M. Every boundary component T of M is an oriented torus, which contains two natural unoriented simple closed curves: the boundary $m = T \cap \partial S$ of the section S and the fibre l of the bundle. If oriented, the curves m and l form a basis (m, l) for $H_1(T, \mathbb{Z})$. We choose orientations for m and l such that (m, l) form a positively oriented basis: there is a unique choice up to reversing both m and l.

A Dehn filling on T is determined as usual by a pair (p, q) of coprime integers that indicate the slope $\pm(pm + ql)$ to be killed.

Suppose now that S has only one boundary component and let M^{fill} be obtained by Dehn filling M along the slope $(1, q)$. Let $\hat{S}$ be the closed surface obtained by capping S with a disc.

Proposition 10.2.5. *The circle bundle $M \to S$ extends to a circle bundle $M^{\text{fill}} \to \hat{S}$. Every oriented circle bundle on $\hat{S}$ is obtained in this way, and distinct values of q yield vector bundles that are not orientation-preservingly isomorphic.*

Proof. The meridian of the filling solid torus is $m' = m + ql$. The fibre l has geometric intersection 1 with m' and is hence a longitude for the filling solid torus. We may represent the filling solid torus as $D \times S^1$ with $m' = S^1 \times \{y\}$ and $l = \{x\} \times S^1$. The circle bundle $M \to S$ extends naturally to a circle bundle $M^{\text{fill}} \to \hat{S}$ with $\hat{S} = S \cup D$.

Every closed circle bundle $N \to \hat{S}$ arises in this way: the bundle above a disc $D \subset \hat{S}$ is the trivial $D \times S^1$, and if we remove it we get $M \to S$ back. The number q is intrinsically determined: the meridian m does not depend on the section of $M \to S$ by Corollary 10.2.3, and the equality $m' = m + ql$ determines q. Therefore distinct values of q yield non-isomorphic bundles. $\square$

The integer q is the *Euler number* of the circle bundle and is usually denoted with the letter e. We summarise our discovery:

Corollary 10.2.6. *For every $e \in \mathbb{Z}$ and every closed surface S there is a unique oriented circle bundle over S with Euler number e.*

A change of orientation for M transforms e into $-e$. Recall that every closed surface S has a trivial circle bundle $S \mathbin{\widetilde{\times}} S^1$ constructed by doubling the unique oriented line bundle on S.

Exercise 10.2.7. An oriented circle bundle over a closed surface is trivial $\Longleftrightarrow e = 0 \Longleftrightarrow$ the bundle has a section.

We may see the Euler number of a bundle $M \to S$ over a closed S as an obstruction for the existence of a section.

Remark 10.2.8. Every oriented n-dimensional vector bundle $E \to S$ over a closed oriented n-manifold S has a *Euler number* defined by taking two generic sections and counting their signed intersections. We briefly explain how this number is closely related to the one we defined here.

Each vector bundle $E \to S$ induces a sphere bundle $M \to S$: it suffices to fix a Riemannian metric on E and take the sub-bundle consisting of unit tangent vectors. When $n = 2$ we get a circle bundle $M \to S$ and the Euler number of $E \to S$ coincides with that of $M \to S$ that we defined above.

When E is the tangent bundle of S, the Euler number is the Euler characteristic $\chi(S)$. For instance, the unit tangent bundle of S^2 has Euler number $e = \chi(S^2) = 2$ and hence it is diffeomorphic to $L(2,1) = \mathbb{RP}^3$.

Exercise 10.2.9. Let M be a circle bundle over the genus-g surface S_g with Euler number e. We have $H_1(M, \mathbb{Z}) = \mathbb{Z}^{2g} \times \mathbb{Z}/_{e\mathbb{Z}}$.

Corollary 10.2.10. *Let $M \to S_g$ and $M' \to S_{g'}$ be circle bundles with Euler numbers e and e'. The manifolds M and M' are diffeomorphic $\Longleftrightarrow g = g'$ and $|e| = |e'|$.*

Exercise 10.2.11. The circle bundle M over S^2 with Euler number e is diffeomorphic to the lens space $L(|e|, 1)$.

Hint. The base sphere S^2 decomposes into two discs, and the fibration over each disc is a solid torus. So M is the union of two solid tori. $\square$

We end the discussion by defining explicitly a famous non-trivial circle bundle over S^2.

10.2.4. The Hopf fibration. The quotient map
$$(\mathbb{C}^2)^* \longrightarrow \mathbb{CP}^1$$
$$(z, w) \longmapsto [z, w]$$
restricts to a circle bundle $S^3 \to \mathbb{CP}^1 = S^2$ called the *Hopf fibration*. The fibre over $[z, w]$ is the circle $(\omega z, \omega w)$ as $\omega \in S^1$ varies. See Figure 10.9.

Exercise 10.2.11 implies that the Euler number of the Hopf fibration $S^3 \to S^2$ is ± 1, and hence it has no section by Exercise 10.2.7.
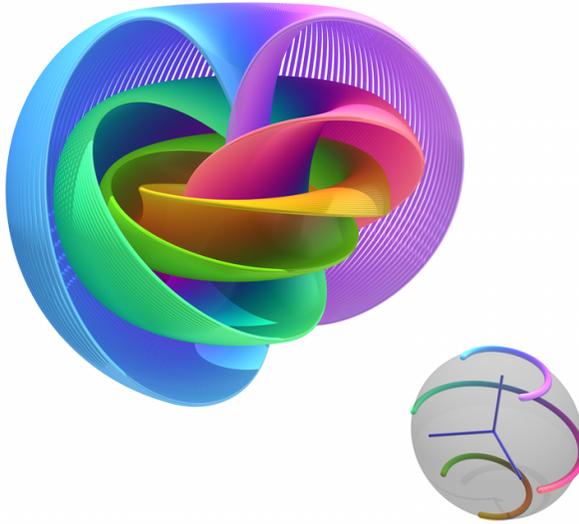
Figure 10.9. The fibre of every point in the Hopf fibration
$S^3 \rightarrow S^2$ is a circle, and the counterimage of a circle in
S^2 is a torus in S^3. The picture shows the counterimage
of three circle arcs: each is a portion of standard torus in
$S^3 = \mathbb{R}^3 \cup \infty$, foliated by circles of type $(1, 1)$.

10.3. Seifert manifolds

We now enlarge the class of circle bundles over surfaces by admitting
some kind of singular fibres. We introduce the *Seifert fibrations*, whose
total spaces are called *Seifert manifolds*. These manifolds were introduced
by Herbert Seifert in 1933 and fit perfectly in the much more recent ge-
ometrisation perspective: in the next chapter we will introduce the eight
three-dimensional geometries, and prove that six of them are realised pre-
cisely by the Seifert manifolds.

10.3.1. Definition. We define the Seifert manifolds as Dehn fillings
of trivial bundles over surfaces with boundary. Here are the details.

Let M be the (unique) oriented bundle $S \mathbin{\tilde{\times}} S^1$ over a compact con-
nected (possibly non-orientable) surface S with boundary. We denote by
S the zero-section.

Let $T_1, \ldots, T_k$ be the boundary tori of M. On each T_i we choose an
orientation for the meridian $m_i = T_i \cap \partial S$ and for the fibre l_i of the bundle
so that the basis (m_i, l_i) for $H_1(T_i, \mathbb{Z})$ be positively oriented.

A (p_i, q_i)-Dehn filling on T_i kills the slope $p_i m_i + q_i l_i$. We say that the
Dehn filling is *fibre-parallel* if $p_i = 0$, *i.e.* if it kills a fibre.

Definition 10.3.1. A *Seifert manifold* is any 3-manifold N obtained from M by Dehn filling some $h \leqslant k$ boundary tori in a non-fibre-parallel way, that is with $p_i \neq 0$ for all i.

The Seifert manifold is closed if $h = k$, and has $k - h$ boundary tori otherwise. It is not important to know which h tori are filled, in virtue of the following.

Proposition 10.3.2. *Every permutation of the boundary tori is realised by a self-diffeomorphism of M that preserves the pairs $\pm(m_i, l_i)$.*

Proof. Every permutation of the boundary circles of S is realised by a self-diffeomorphism of S, that extends orientation-preservingly to the orientable I-bundle and its double M. □

The pair (p_i, q_i) is determined up to sign, so we can always suppose $p_i > 0$ and we fully encode the Seifert manifold N using the following notation:

$$(6) \qquad N = \big(\hat{S}, (p_1, q_1), \ldots, (p_h, q_h)\big)$$

where $\hat{S}$ is S with h boundary components capped. The reason for using $\hat{S}$ instead of S is that N has a particular fibration onto $\hat{S}$, as we will soon see. Before constructing this fibration we list some simple examples that should hopefully help the reader to familiarise with the notation (6), that will be used extensively in the whole chapter.

Example 10.3.3. The Seifert manifold $\big(S_g, (1, e)\big)$ is the circle bundle over the orientable genus-g surface S_g with Euler number e, by construction. In particular $\big(S_g, (1, 0)\big) = S_g \times S^1$.

Example 10.3.4. The Seifert manifold $\big(S^2, (p, q)\big)$ is diffeomorphic to the lens space $L(q, p)$.

The following facts follow from Exercise 10.1.7.

Exercise 10.3.5. The Seifert manifold $\big(D^2, (p, q)\big)$ is a solid torus.

In the following exercise, it should be clear that the resulting manifold is a lens space (by Exercise 10.1.7) and the hard part is to check carefully that the proposed lens space parameters are correct.

Exercise 10.3.6. The Seifert manifold $\big(S^2, (p_1, q_1), (p_2, q_2)\big)$ is the lens space $L(p_1 q_2 + q_1 p_2, r q_2 + s p_2)$ where r, s are such that $p_1 s - q_1 r = \pm 1$. In particular, this Seifert manifold is S^3 when $p_1 q_2 + q_1 p_2 = \pm 1$.

10.3.2. Seifert fibrations. As we anticipated, a Seifert manifold N as in (6) possesses some kind of singular fibration over the filled surface $\hat{S}$. We clarify this point here by defining the notion of *Seifert fibration*.

Let (p, q) be two coprime integers with $p > 0$. A *standard fibered solid torus* with coefficients (p, q) is the solid torus
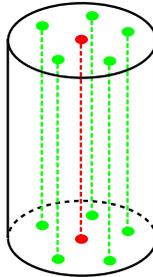
$$D \times [0, 1]/_\psi$$

Figure 10.10. A standard fibered solid torus. We identify the top and bottom discs by a $2\pi\frac{q}{p}$ rotation, for some q coprime with p. Here $p = 5$. Every non-central fibre (green) winds p times along the central fibre (red).

where $\psi\colon D \times 0 \to D \times 1$ is a rotation of angle $2\pi\frac{q}{p}$. The fibration into vertical segments $\{\mathrm{pt}\} \times [0, 1]$ extends to a fibration into circles of the solid torus. The central fibre obtained by identifying the endpoints of $0 \times [0, 1]$ is the core of the solid torus, and every non-central fibre winds p times around the core of M: see Figure 10.10.

The positive number p is the *multiplicity* of the central fibre. If $p = 1$ the fibered solid torus is diffeomorphic to the usual product fibration $D \times S^1$ and the central fibre is *regular*. If $p > 1$ the central fibre is *singular*.

Definition 10.3.7. A *Seifert fibration* is a partition of a compact oriented 3-manifold N with (possibly empty) boundary into circles, such that every circle has a fibered neighbourhood diffeomorphic to a standard fibered solid torus.

Let S be the topological space obtained from N by quotienting circles to points.

Proposition 10.3.8. *The space S is a compact connected surface with (possibly empty) boundary.*

Proof. Every standard fibered solid torus quotients to a disc: a sector $z = \rho e^{i\theta}$ with $0 \leqslant \rho \leqslant 1$ and $0 \leqslant \theta \leqslant \frac{2\pi}{p}$ with two edges $\theta = 0$ and $\theta = \frac{2\pi}{p}$ identified by a rotation. This disc actually looks like a "cone", and this motivates the following discussion.   □

The map $N \to S$ is in fact what we call a Seifert fibration. The surface S may have boundary and may be non-orientable, and its interior has a natural orbifold structure: if the preimage of $x \in S$ is a fibre of order p, we see x as a cone point of order p, see Section 6.2.4. If N has boundary, then S also has, and we say that S itself is an orbifold for simplicity although we actually mean only its interior. Morally, we should consider the fibration $N \to S$ as a circle bundle over the orbifold S.

A Seifert fibration without singular fibres is just an ordinary circle bundle. We now show that Seifert manifolds and Seifert fibrations are more or less the same thing.

Proposition 10.3.9. *The Seifert manifold*

$$N = \big(S, (p_1, q_1), \dots, (p_h, q_h)\big)$$

has a Seifert fibration $N \to S$ over the orbifold

$$(S, p_1, \dots, p_h).$$

Every Seifert fibration arises in this way.

Proof. The Seifert manifold N is obtained by filling h components of a bundle $M = S' \stackrel{(\sim)}{\times} S^1$. The bundle $M \to S'$ extends to a Seifert fibration $N \to S$ where the orbifold S is obtained from S' by attaching h discs with cone points $p_1, \dots, p_h$.

More precisely, we fill each boundary torus T_i of M with a solid torus having meridian $\mu_i = p_i m_i + q_i l_i$. We fix a longitude $\lambda_i = r_i m_i + s_i l_i$ for this solid torus by choosing r_i, s_i with $p_i s_i - q_i r_i = 1$. We get $l_i = p_i \lambda_i - r_i \mu_i$. By hypothesis the Dehn filling is not fibre-parallel, hence $p_i \neq 0$ and the fibration $M \to S'$ extends to a standard fibration of the solid torus with coefficients $(p_i, -r_i)$.

Every Seifert fibration $N \to S$ arises in this way: if we remove the singular fibres (or a regular one, if there are not) we get an ordinary circle bundle over a surface with boundary, which is trivial. Therefore N is a Dehn filling of this trivial bundle, hence a Seifert manifold. $\qquad\square$

We have seen that the notation

(7) $$N = \big(S, (p_1, q_1), \dots, (p_h, q_h)\big)$$

defines a Seifert fibration $N \to S$ and a Seifert manifold N.

Example 10.3.10. If the orbifold S is a disc with at most one singular point then N is a standard fibered solid torus. If S is S^2 with at most 2 singular points then N is a lens space: see Exercises 10.3.5 and 10.3.6.

10.3.3. Classification of Seifert fibrations. We say that two Seifert fibrations $\pi_1 \colon N_1 \to S$, $\pi_2 \colon N_2 \to S$ are *isomorphic* if there is a diffeomorphism $\psi \colon N_1 \to N_2$ such that $\pi_1 = \pi_2 \circ \psi$. Two different notations as in (7) may describe isomorphic fibrations, but this phenomenon is completely understood.

Proposition 10.3.11. *Two notations as in (7) describe two orientation-preservingly isomorphic Seifert fibrations if and only if they are related by a finite sequence of the following moves and their inverses:*

(8) $$(p_i, q_i), (p_{i+1}, q_{i+1}) \longmapsto (p_i, q_i + p_i), (p_{i+1}, q_{i+1} - p_{i+1}),$$

(9) $$(p_1, q_1), \dots, (p_h, q_h) \longmapsto (p_1, q_1), \dots, (p_h, q_h), (1, 0),$$

(10) $$(p_i, q_i) \longmapsto (p_i, q_i + p_i) \quad \text{if } \partial N \neq \varnothing,$$

and permutations of the pairs (p_i, q_i)'s.

Proof. Recall that N is a Dehn filling of $M = S' \mathbin{\widetilde{\times}} S^1$. Move (8) is the result of twisting M along a fibered annulus A connecting the tori T_i and T_{i+1} in ∂M: this self-diffeomorphism of M acts on T_i and T_{i+1} like two opposite Dehn twists and extends to an isomorphism of the two fibrations. In (10) we twist along an annulus connecting T_i and ∂N. The move (9) corresponds to drilling a nonsingular fibered torus and refilling it back.

We now prove that these moves suffice to connect two isomorphic Seifert fibrations. Suppose that two distinct notations as in (7) describe isomorphic fibrations. We use the moves (8), (9), and (10) to eliminate the parameters $p_i = 1$ as much as possible from both notations. If there are no singular fibres, we end up with a single parameter $(1, e)$ if S is closed, and no parameters at all if $\partial S \neq \varnothing$. We conclude by Corollary 10.2.6.

If the fibration has at least one singular fibre, we can eliminate all $p_i = 1$ and the (p_i, q_i) correspond to singular fibres. An isomorphism of Seifert fibrations sends singular fibres to singular fibres and hence induces an isomorphism of their complement $S' \mathbin{\widetilde{\times}} S^1$. The parameters (p_i, q_i) are determined by the choice of a section in $S' \mathbin{\widetilde{\times}} S^1$. Different sections are related by Dehn twist along annuli and hence the parameters are related by the moves (8) and (10). $\qquad\square$

Proposition 10.3.11 classifies all the Seifert fibrations up to isomorphism. A classification of Seifert manifolds up to *diffeomorphism* would also be desirable, but it is much harder to obtain because a three-manifold may admit many non-isomorphic Seifert fibrations. For instance, Exercise 10.3.6 shows that the lens spaces may fibre in many different ways; a manifold as familiar as S^3 fibres over the orbifold (S^2, p_1, p_2) if p_1 and p_2 are coprime and hence has infinitely many non-isomorphic fibrations. It is a stimulating exercise to try to visualise these Seifert fibrations of S^3.

We now start a long journey in Seifert manifolds theory, whose ultimate goal is to classify them completely up to diffeomorphism. We will see at the end that the only Seifert manifolds admitting non-isomorphic fibrations are the "smallest" ones, like S^3, the lens spaces, and few more that will be classified using some *ad hoc* argument.

Exercise 10.3.12. The number of non-isomorphic Seifert fibrations over a fixed orbifold S is finite $\Longleftrightarrow \partial S \neq \varnothing$.

10.3.4. Euler number. We now extend the notion of Euler number from ordinary to Seifert fibrations. We define the *Euler number* of the fibration (7) to be the rational number

$$e(N) = \sum_{i=1}^{h} \frac{q_i}{p_i}.$$

The Euler number is only defined modulo $\mathbb{Z}$ when N has boundary. The good definition follows from Proposition 10.3.11 (the moves do not affect e, except (10) that modifies e into $e + 1$ and applies only when N has boundary) and is coherent with the circle bundle case. The Euler number depends on the fibration and not only on N, but we write $e(N)$ anyway for simplicity. Proposition 10.3.11 easily implies the following.

Corollary 10.3.13. *Two Seifert fibrations*

$$\big(S, (p_1, q_1), \ldots, (p_h, q_h)\big), \quad \big(S', (p_1', q_1'), \ldots, (p_{h'}', q_{h'}')\big)$$

with $p_i, p_i' \geqslant 2$ are orientation-preservingly isomorphic if and only if $S = S'$, $h = h'$, $e = e'$, and up to reordering $p_i = p_i'$ and $q_i \equiv q_i'$ (mod p_i) for all i.

The numbers e and e' indicate the Euler numbers of the two fibrations, and recall that they are only defined modulo $\mathbb{Z}$ when $\partial S \neq \varnothing$.

Remark 10.3.14. The move

$$\big(S, (p_1, q_1), \ldots, (p_h, q_h)\big) \longmapsto \big(S, (p_1, -q_1), \ldots, (p_h, -q_h)\big)$$

corresponds to a change of orientation for the three-manifold and transforms e into $-e$.

10.3.5. Homology. We now characterise the Seifert manifolds that have finite homology groups, and in particular the homology spheres.

Proposition 10.3.15. *The homology group $H_1(M, \mathbb{Z})$ of*

$$M = \big(S, (p_1, q_1), \ldots, (p_h, q_h)\big)$$

is finite $\Longleftrightarrow$ one of the following holds:

- *$S = S^2$ and $e \neq 0$, and we get $|H_1(M, \mathbb{Z})| = |e|p_1 \cdots p_h$;*
- *$S = \mathbb{RP}^2$, and we get $|H_1(M, \mathbb{Z})| = 4p_1 \cdots p_h$.*

Proof. If M is not closed then $H_1(M, \mathbb{Z})$ is infinite because $H_1(\partial M, \mathbb{Z})$ is. If $S \neq S^2, \mathbb{RP}^2$ then $H_1(S)$ is infinite. It is easy to check that the fibration $M \to S$ induces a surjection $H_1(M) \to H_1(S)$ and hence $H_1(M)$ is also infinite.

Suppose $S = S^2$. The manifold M is a (p_i, q_i)-Dehn filling of $S_{0,h} \times S^1$ where $S_{0,h}$ is the sphere with h holes. The homology of $S_{0,h} \times S^1$ is generated by $m_1, \ldots, m_h, l$ with the relation $m_1 + \ldots + m_h = 0$. The i-th Dehn filling adds the relation $p_i m_i + q_i l_i = 0$. The relations form a square $(h + 1)$-matrix

$$\begin{pmatrix} 1 & p_1 & 0 & \cdots & 0 \\ 1 & 0 & p_2 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 1 & 0 & 0 & \cdots & p_h \\ 0 & q_1 & q_2 & \cdots & q_h \end{pmatrix} \sim \begin{pmatrix} 1 & p_1 & 0 & \cdots & 0 \\ 1 & 0 & p_2 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 1 & 0 & 0 & \cdots & p_h \\ -e & 0 & 0 & \cdots & 0 \end{pmatrix}.$$

We have used Gauss moves to simplify the last row. The determinant of this matrix is $\pm e p_1 \cdots p_h$. The order $|H_1(M, \mathbb{Z})|$ is the absolute value of the determinant if it is non-zero, and is infinite if it is zero.

Suppose $S = \mathbb{RP}^2$ and let $S_{1,h}^{no}$ be $\mathbb{RP}^2$ with h holes. The homology of $S_{1,h}^{no} \mathbin{\widetilde{\times}} S^1$ is generated by $a, m_1, \ldots, m_h, l$, where a is an orientation-reversing curve in $S_{1,h}^{no}$, with the relations $2a + m_1 + \ldots + m_h = 0$ and $2l = 0$. The relations now form a square $(h + 2)$-matrix with determinant

$$\det \begin{pmatrix} 2 & 0 & 0 & \ldots & 0 \\ 1 & 0 & p_1 & \ddots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \cdots & p_h \\ 0 & 2 & q_1 & \cdots & q_h \end{pmatrix} = \pm 2 \det \begin{pmatrix} 2 & 0 & \cdots & 0 \\ 1 & p_1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & \cdots & p_h \end{pmatrix} = \pm 4 p_1 \cdots p_h.$$

The proof is complete. $\qquad\square$

We deduce an elegant description of all the Seifert homology spheres.

Corollary 10.3.16. *For every set $p_1, \ldots, p_h$ of pairwise coprime integers $p_i \geqslant 2$ there is a unique homology sphere $\Sigma(p_1, \ldots, p_h)$ Seifert-fibering over $(S^2, p_1, \ldots, p_h)$. Every Seifert homology sphere arises in this way.*

Proof. By the previous proposition a Seifert manifold

$$M = \big(S, (p_1, q_1), \ldots, (p_h, q_h)\big)$$

is a homology sphere if and only if $S = S^2$ and $|e p_1 \cdots p_h| = 1$. We have

$$e p_1 \cdots p_h = \sum_{i=1}^{h} \frac{q_i}{p_i} p_1 \cdots p_h = \sum_{i=1}^{h} q_i p_1 \cdots \widehat{p_i} \cdots p_h = \sum_{i=1}^{h} q_i p_i'$$

where we set $p_i' = p_1 \cdots \widehat{p_i} \cdots p_h$. The integers $p_1, \ldots, p_h$ are *pairwise* coprime if and only if $p_1', \ldots, p_h'$ are *globally* coprime (no prime number divides all of them). The equation $\sum_{i=1} q_i p_i' = 1$ is satisfied by some $q_1, \ldots, q_h \Longleftrightarrow$ they are globally coprime. Different solutions q_i are related by moves as in Proposition 10.3.11 (exercise) and produce the same Seifert fibration. $\qquad\square$

For instance, the homology spheres $\Sigma(p_1)$ and $\Sigma(p_1, p_2)$ are just S^3 by Example 10.3.10. The simplest homology sphere with three singular fibres is $\Sigma(2, 3, 5)$. This manifold is called the *Poincaré homology sphere* and we will soon see that its fundamental group has order 120. All the other Seifert homology spheres have infinite fundamental group: this is related to the fact that $\sum \frac{1}{p_i} < h - 2$ for all choices of $(p_1, \ldots, p_h)$ with $h \geqslant 3$ except $(2, 3, 5)$, as we will soon see. A Seifert homology sphere of type $\Sigma(p_1, p_2, p_3)$ is called a *Brieskorn homology sphere*.

10.3.6. Coverings. We now start to investigate the coverings of Seifert fibered spaces. Our main goal will be to subdivide the Seifert manifolds into nine classes: this will be done in Section 10.3.9.

Like ordinary fibrations, Seifert fibrations behave well with respect to coverings. Let $M \to S$ be a Seifert fibration and $\tilde{M} \to M$ a covering. The foliation into circles of M lifts to a foliation into circles or lines in $\tilde{M}$, with some quotient space $\tilde{S}$.

Proposition 10.3.17. *The quotient $\tilde{S}$ is an orbifold covering of S.*
- *If $\tilde{M}$ foliates in circles then $\tilde{M} \to \tilde{S}$ is a Seifert fibration,*
- *If $\tilde{M}$ foliates in lines then $\tilde{M} \to \tilde{S}$ is a line bundle.*

In the second case $\tilde{S}$ has no singular points.

Proof. This holds on all coverings of a standard fibered solid torus and hence holds everywhere. □

We now concentrate ourselves on the finite-degree case.

10.3.7. Finite-degree coverings. We define a *finite-degree covering* of a Seifert fibration $M \to S$ to be a commutative diagram

$$\begin{array}{ccc} \tilde{M} & \longrightarrow & M \\ \downarrow & & \downarrow \\ \tilde{S} & \longrightarrow & S \end{array}$$

where $\tilde{M} \to \tilde{S}$ is a Seifert fibration, $\tilde{M} \to M$ is a finite-degree covering, and $\tilde{S} \to S$ is an orbifold covering. Proposition 10.3.17 implies the following.

Corollary 10.3.18. *Let $M \to S$ be a Seifert fibration. Every finite-degree covering $\tilde{M} \to M$ induces a finite-degree covering of Seifert fibrations:*

$$\begin{array}{ccc} \tilde{M} & \longrightarrow & M \\ \vdots & & \downarrow \\ \tilde{S} & \cdots\cdots & S \end{array}$$

The dotted arrows indicate the maps that are induced. The degree d of such a covering $\tilde{M} \to M$ splits into two parts:
- the *horizontal* degree d_h is the degree of the covering $\tilde{S} \to S$,
- the *vertical* degree d_v is the degree with which a regular fibre of $\tilde{M}$ covers a regular fibre of M.

The vertical degree is well-defined since regular fibres in $\tilde{M}$ form a connected set. The pre-image of a regular fibre in M consists of d_h regular fibres in $\tilde{M}$, each fibering with degree d_v. Therefore

$$d = d_h \cdot d_v.$$

A covering $\tilde{M} \to M$ is *horizontal* or *vertical* if respectively $d_v = 1$ or $d_h = 1$.

Proposition 10.3.19 (Pull-back). *Let $M \to S$ be a Seifert fibration. Every finite-degree orbifold covering $\tilde{S} \to S$ is induced by a unique horizontal covering of Seifert fibrations:*

$$
\begin{array}{ccc}
\tilde{M} & \dashrightarrow & M \\
\downarrow & & \downarrow \\
\tilde{S} & \longrightarrow & S
\end{array}
$$

Proof. Like for ordinary bundles, there is a unique way to define $\tilde{M}$ by pulling back the Seifert fibration on fibered solid tori. □

Recall that an orbifold is *very good* when it is finitely covered by a surface. Every locally orientable 2-orbifold is very good except the bad orbifolds $S^2(p_1)$ and $S^2(p_1, p_2)$ with $p_1 \neq p_2$, see Theorem 6.2.10 and Corollary 6.2.11.

Corollary 10.3.20. *If S is good, every Seifert fibration $M \to S$ is finitely covered by a circle bundle over a surface.*

Proof. Pull-back the fibration along the surface cover $\tilde{S} \to S$. □

Proposition 10.3.21. *Every finite-degree covering between Seifert fibrations is a composition of one vertical and one horizontal covering.*

Proof. Let $p \colon \tilde{M} \to M$ be a covering of Seifert fibrations, with base spaces $\tilde{S} \to S$. If we pull-back M to $\tilde{S}$ we get a horizontal covering $p_h \colon M_h \to M$. There is a natural vertical $p_v \colon \tilde{M} \to M_h$ such that $p = p_h \circ p_v$. □

10.3.8. Euler number and Euler characteristic. The Euler number of the fibration and the Euler characteristic of the base orbifold (defined in Section 6.2.4) behave well with coverings.

Proposition 10.3.22. *Let $p \colon \tilde{M} \to M$ be a finite covering of closed Seifert fibrations with base orbifolds $\tilde{S}$, S, with degrees (d_h, d_v). We get*

$$\chi(\tilde{S}) = d_h \cdot \chi(S),$$

$$e(\tilde{M}) = \frac{d_h}{d_v} \cdot e(M).$$

Proof. The first equality holds for every orbifold covering $\tilde{S} \to S$. Concerning the second one, by Proposition 10.3.21 we may suppose p is either vertical or horizontal. Write

$$M = \big(S, (p_1, q_1), \ldots (p_h, q_h)\big)$$

and recall that M is obtained from a circle bundle N by (p_i, q_i)-filling the boundary torus T_i for all $i = 1, \ldots, h$. Set $\tilde{T}_i = p^{-1}(T_i)$ and $\tilde{N} = p^{-1}(N)$.

If p is vertical, we fix a section of $\tilde{N}$ and note that it projects to a section of N: these sections induce meridians on $\tilde{T}_i$ and T_i. Here $\tilde{T}_i$ is a

single torus and $\tilde{T}_i \to T_i$ wraps the fibre with degree $d = d_v$. Since the covering extends to the filled solid tori, we have $d | q_i$ and $\tilde{T}_i$ is filled with parameters $(p_i, q_i/d)$. Therefore $e(\tilde{M}) = e(M)/d$.

If p is horizontal, we fix a section of N, and its counterimage is a section of $\tilde{N}$: these sections induce meridians on T_i and $\tilde{T}_i$. Now $\tilde{T}_i$ consists of some tori $\tilde{T}_i = \tilde{T}_{i,1} \sqcup \ldots \sqcup \tilde{T}_{i,k_i}$ and $\tilde{T}_{i,j} \to T_i$ wraps the meridians with some degree $d_{i,j}$. The total sum of these local degrees is $d_{i,1} + \ldots + d_{i,k_i} = d = d_h$. Similarly as above we have $d_{i,j} | p_i$ and the filling solid torus at $\tilde{T}_{i,j}$ has parameters $(p_i/d_{i,j}, q_i)$. Therefore

$$e(\tilde{M}) = \sum_{i=1}^{h} \sum_{j=1}^{k_i} \frac{d_{i,j} q_i}{p_i} = \sum_{i=1}^{h} d \cdot \frac{q_i}{p_i} = d \cdot e(M).$$

This concludes the proof. □

An important consequence is that the signs of $\chi(S)$ and $e(M)$ are invariant under finite coverings.

10.3.9. Commensurability classes. We now would like to subdivide the Seifert manifolds into few classes, and to this purpose we introduce a general equivalence relation between manifolds.

Definition 10.3.23. Two manifolds M and N are *commensurable* if there is a manifold that covers both M and N with finite degrees.

Proposition 10.3.24. *Commensurability is an equivalence relation.*

Proof. If M is commensurable with N_1 and N_2, it has finite-sheeted coverings M_1, M_2 that cover N_1 and N_2 corresponding to finite-index subgroups $\Gamma_1, \Gamma_2 < \pi_1(M)$. The subgroup $\Gamma = \Gamma_1 \cap \Gamma_2$ has also finite index and determines a manifold that covers both N_1 and N_2. □

Exercise 10.3.25. There are three commensurable classes of closed surfaces, determined by their Euler characteristic being positive, null, or negative.

We now classify the commensurability classes of Seifert manifolds. The *3-torus* is of course $T \times S^1 = S^1 \times S^1 \times S^1$.

Proposition 10.3.26. *A closed Seifert fibration $M \to S$ has:*
- *$\chi(S) > 0$ and $e = 0 \iff M$ is covered by $S^2 \times S^1$,*
- *$\chi(S) > 0$ and $e \neq 0 \iff M$ is covered by S^3,*
- *$\chi(S) = 0$ and $e = 0 \iff M$ is covered by the 3-torus,*
- *$\chi(S) = 0$ and $e \neq 0 \iff M$ is covered by a twisted bundle over T,*
- *$\chi(S) < 0$ and $e = 0 \iff M$ is covered by $S_g \times S^1$ for some $g > 1$,*
- *$\chi(S) < 0$ and $e \neq 0 \iff M$ is covered by a twisted bundle over S_g for some $g > 1$.*

	$\chi > 0$	$\chi = 0$	$\chi < 0$
$e = 0$	$S^2 \times S^1$	$S^1 \times S^1 \times S^1$	$S_2 \times S^1$
$e \neq 0$	S^3	$(T, (1,1))$	$(S_2, (1,1))$
$\partial M \neq \varnothing$	$D \times S^1$	$T \times [0,1]$	$P \times S^1$

Table 10.1. There are 9 commensurability classes of Seifert manifolds: 6 closed and 3 with boundary. Every Seifert manifold M is commensurable with one (and only one) of these 9 manifolds. The commensurability class of M is easily detected by looking at its invariants χ and e (the latter only in the closed case). The surfaces T, S_2, D, P are the torus, the genus-2 surface, the disc, and the pair-of-pants. Note that $S^3 = (S^2, (1,1))$.

Proof. If S is a bad orbifold, then $S = S^2(p)$ or $S^2(p_1, p_2)$ with $p_1 \neq p_2$ and hence we get $e \neq 0$ and $\chi(S) > 0$. The manifold M is a lens space by Example 10.3.10 and is hence covered by S^3.

If S is good, the fibration $M \to S$ is covered by a circle bundle $\tilde{M} \to \tilde{S}$ over an orientable closed surface $\tilde{S}$ by Corollary 10.3.20. By Proposition 10.3.22 the numbers $\chi(\tilde{S})$ and $e(\tilde{M})$ have the same signs of $\chi(S)$ and $e(M)$. Therefore $\tilde{S} = S^2, T$, or S_g with $g > 1$, depending on whether $\chi(S)$ is positive, null, or negative. Exercise 10.2.7 says that the circle bundle $\tilde{M} \to \tilde{S}$ is trivial $\Longleftrightarrow e(\tilde{M}) = 0 \Longleftrightarrow e(M) = 0$. Note that a non-trivial bundle over S^2 is a lens space $L(e, 1)$ with $e \neq 0$ and is hence covered by S^3.

We have proved that M is covered (according to the signs of χ and e) by a manifold belonging to one of the six types: $S^3, S^2 \times S^1$, the 3-torus, a twisted bundle over $T, S_g \times S^1$, and a twisted bundle over S_g. It remains to prove that M cannot be covered by two manifolds M_1, M_2 belonging to two different types: this holds because manifolds of distinct types are not commensurable. To prove that, note that the finite cover of a manifold of one of the six types is a manifold of the same type, and Corollary 10.2.10 implies that a manifold cannot belong to two different types. If two manifolds of distinct types were commensurable they would be covered by a manifold belonging to both types, yielding a contradiction. □

We can now easily classify all closed Seifert manifolds up to commensurability. To conclude we just need to solve the following exercise.

Exercise 10.3.27. Pick $S = S_g$ and $e > 0$. Construct:

- a degree-e vertical covering $(S, (1,1)) \to (S, (1,e))$,
- a degree-e horizontal covering $(\tilde{S}, (1,e)) \to (S, (1,1))$ if $g \geqslant 1$.

Hint. If $g \geqslant 1$ then S has covers of any degree e. To construct them, pick surjective homomorphisms $\pi_1(S) \to H_1(S) \to \mathbb{Z}/_{e\mathbb{Z}}$. $\qquad$ $\square$

Corollary 10.3.28. *There are six commensurability classes of closed Seifert manifolds, depending on χ and e as shown in Table 10.1.*

Proof. By the previous exercise the non-trivial (or trivial) bundles over closed surfaces with $\chi < 0$ (or $\chi = 0$, $\chi > 0$) are all commensurable. $\square$

Corollary 10.3.29. *A closed Seifert fibration $M \to S$ has $e = 0 \iff$ it is finitely covered by a trivial circle bundle.*

We recall that a Seifert manifold M may have non-isomorphic Seifert fibrations $M \to S$ and $M \to S'$, and the invariants χ, χ' and e, e' of the two fibrations are not necessarily equal; however χ is positive, null, or negative if and only if χ' is, and e vanishes if and only if e' does (in the closed case). This holds because the commensurability class of M does not depend on its fibration. For instance, Exercise 10.3.6 shows that a lens space $M = L(p, q)$ with $p > 0$ has many fibrations, but each with $\chi > 0$ and $e \neq 0$.

We now consider the boundary case, which is simpler because circle bundles over orientable surfaces with boundary are always trivial. Let A denote the annulus and $S_{g,b}$ be the surface of genus g with b open discs removed.

Proposition 10.3.30. *A Seifert fibration $M \to S$ with boundary has*

- *$\chi(S) > 0 \iff M = D \times S^1$,*
- *$\chi(S) = 0 \iff M$ is covered by $A \times S^1 = T \times [0, 1]$,*
- *$\chi(S) < 0 \iff M$ is covered by $S_{g,b} \times S^1$ for some $g + b > 2$.*

Proof. Every Seifert manifold with boundary is covered by a circle bundle over an orientable surface, and such a bundle is trivial here. If $\chi > 0$ the base surface is a disc with at most one cone point and hence M is a standard fibered solid torus. $\square$

Corollary 10.3.31. *There are three commensurability classes of Seifert manifolds with boundary, depending on χ as shown in Table 10.1.*

Proof. The surfaces $S_{g,b}$, with $b > 0$ and $g + b > 2$ are commensurable (exercise). $\square$

We now characterise some commensurability classes by looking at the fundamental groups.

	$\chi > 0$	$\chi = 0$	$\chi < 0$
$e = 0$	$\mathbb{Z}$	$\mathbb{Z}^3$	
$e \neq 0$	$\{e\}$		
$\partial M \neq \varnothing$	$\mathbb{Z}$	$\mathbb{Z}^2$	

Table 10.2. There are five Seifert manifolds M with $\pi_1(M) = \mathbb{Z}^h$ and they belong to distinct commensurability classes. The fundamental group $\pi_1(N)$ of every other manifold N in these five classes is virtually abelian: it contains $\mathbb{Z}^h$ as a finite-index subgroup.

10.3.10. Virtually abelian fundamental groups. A group G is *virtually abelian of rank h* if it contains $\mathbb{Z}^h$ as a finite-index subgroup.

This is a finite-index-independent property: if $G' < G$ has finite index, the group G' is virtually abelian of rank h if and only if G is. Therefore a manifold M has a virtually abelian fundamental group of rank $h \Longleftrightarrow$ every manifold N commensurable with M also has $\Longleftrightarrow$ there is a manifold N in the commensurability class with $\pi_1(N) = \mathbb{Z}^h$.

Proposition 10.3.32. *There are five commensurability classes of Seifert manifolds with virtually abelian fundamental groups, shown in Table 10.2.*

Proof. These five classes contain $S^2 \times S^1$, $S^1 \times S^1 \times S^1$, S^3, $D \times S^1$, and $A \times S^1$. The remaining four classes do not contain manifolds with abelian fundamental group: every manifold in these classes is covered by some $S \times S^1$ with $\chi(S) < 0$ or $\big(S_g, (1, e)\big)$ with $g \geqslant 1$ and $e \geqslant 1$ (use Exercise 10.3.27). The fundamental group of such manifolds is not $\mathbb{Z}^h$: the former because $\pi_1(S)$ is not abelian, and the latter because its abelianization contains torsion, see Exercise 10.2.9.   □

It could be reasonable to expect these five commensurability classes to contain manifolds of some simpler topological nature, which are easier to study and classify. We will see however in Section 10.4.7 that the Seifert manifolds covered by S^3 are quite interesting and their topological classification is certainly not immediate: we already experienced that for lens spaces in Theorem 10.1.12, whose classification is quite involved. The other four classes are indeed of a simpler nature: we now prove that each contains finitely many manifolds, and we classify them completely.

10.3.11. Finite commensurability classes. We now completely identify the commensurability classes of Seifert manifolds that contain only finitely many manifolds.

We start by considering manifolds M with boundary. We have already seen that the class $\chi > 0$ contains only the solid torus. The class $\chi = 0$ also contains few manifolds:
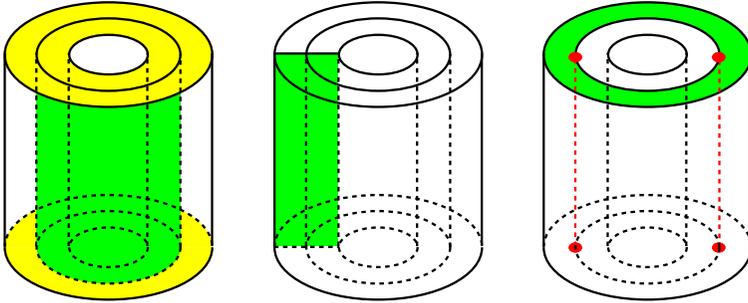
Figure 10.11. Draw the orientable line bundle over the Klein bottle as $S^1 \times [-1, 1] \times [-1, 1]/_\psi$ where ψ identifies the lower and upper (yellow) annuli $S^1 \times [-1, 1] \times \{\pm 1\}$ by gluing $(e^{i\theta}, t)$ to $(e^{-i\theta}, -t)$. The Klein bottle is drawn in green (left). This manifold has two fibrations: a horizontal one with fibres $S^1 \times x \times y$ and base surface the green Möbius strip (centre), and a vertical one with fibres $x \times y \times [-1, 1]$ and base surface the green disc with two cone points of order two (the green annulus becomes a disc after identifying the points $(e^{i\theta}, 0)$ and $(e^{-i\theta}, 0)$). The two singular fibres of order two lie above $(\pm 1, 0)$ and are drawn in red (right).

Proposition 10.3.33. *Every Seifert fibration with boundary and $\chi = 0$ is isomorphic to one of the following:*

$$A \times S^1, \qquad S \mathbin{\widetilde{\times}} S^1, \qquad \big(D, (2, 1), (2, -1)\big).$$

The Seifert manifold is correspondingly diffeomorphic to the interval bundles $T \times I$, $K \mathbin{\widetilde{\times}} I$, and $K \mathbin{\widetilde{\times}} I$ again. Here A, S, and K are the annulus, the Möbius strip, and the Klein bottle.

Proof. The orbifolds with boundary and $\chi = 0$ are A, S, and $(D, 2, 2)$. Every Seifert fibration over $(D, 2, 2)$ is isomorphic to $\big(D, (2, 2n+1), (2, 2m+1)\big)$ and does not depend on $n, m \in \mathbb{Z}$ by Proposition 10.3.11. Hence there is only one Seifert fibration over the orbifold $(D, 2, 2)$, which we write as $\big(D, (2, 1), (2, -1)\big)$ only for aesthetic reasons (to get $e = 0$).

We get three manifolds and we now prove that they are all diffeomorphic to some interval bundle. The diffeomorphism $A \times S^1 \cong T \times I$ is obvious. Draw $K \mathbin{\widetilde{\times}} I$ as $S^1 \times [-1, 1] \times [-1, 1]/_\psi$ as in Figure 10.11. The manifold has two fibrations: a horizontal one by circles $S^1 \times x \times y$, and a vertical one by segments $x \times y \times [-1, 1]$. The horizontal one gives $S \mathbin{\widetilde{\times}} S^1$ and the vertical one closes to a Seifert fibration over the disc with two singular fibres of order two, and hence is $(D, (2, 1), (2, -1))$. □

Corollary 10.3.34. *Every Seifert manifold with boundary and $\chi \geqslant 0$ is diffeomorphic to $D \times S^1$, $T \times I$, or $K \mathbin{\tilde{\times}} I$.*

We now turn to closed Seifert manifolds. We have seen that $K \mathbin{\tilde{\times}} I$ fibres in two non-isomorphic ways, and this has some consequences.

Corollary 10.3.35. *The following diffeomorphisms hold:*
$$\left(S^2, (2, 1), (2, -1), (p, q)\right) \cong \left(\mathbb{RP}^2, (q, p)\right),$$
$$\left(S^2, (2, 1), (2, 1), (2, -1), (2, -1)\right) \cong K \mathbin{\tilde{\times}} S^1.$$

Proof. Both equalities follow from $\left(D, (2, 1), (2, -1)\right) \cong S \mathbin{\tilde{\times}} S^1$. In the first we look at Figure 10.11 to check that a (p, q) curve in $\left(D, (2, 1), (2, -1)\right)$ becomes a (q, p) curve in $S \mathbin{\tilde{\times}} S^1$. The second equality is obtained by doubling the line bundle $K \mathbin{\tilde{\times}} I$.   □

We want to classify the closed Seifert manifolds with $\chi \geqslant 0$ and $e = 0$. We start with the case $\chi > 0$.

Proposition 10.3.36. *Every closed Seifert fibration with $\chi > 0$ and $e = 0$ is isomorphic to one of the following:*
$$S^2 \times S^1, \qquad \mathbb{RP}^2 \mathbin{\tilde{\times}} S^1, \qquad \left(S^2, (p, q), (p, -q)\right).$$
The manifolds of the last type are all diffeomorphic to $S^2 \times S^1$.

Proof. If the base surface S is a sphere with $\leqslant 2$ singular fibres, we use Exercise 10.3.6. Otherwise S is one of the following orbifolds (see Table 6.1):
$$(S^2, 2, 2, p), \quad (S^2, 2, 3, 3), \quad (S^2, 2, 3, 4), \quad (S^2, 2, 3, 5), \quad \mathbb{RP}^2, \quad (\mathbb{RP}^2, p)$$
with $p \geqslant 2$. In all cases except $\mathbb{RP}^2$, we get $e \neq 0$ for any choice of Dehn filling parameters: for instance
$$e\left(S^2, (2, q_1), (2, q_2), (p, q_3)\right) = \frac{q_1 + q_2}{2} + \frac{q_3}{p} \neq 0.$$
The other cases are analogous.   □

The manifold $\mathbb{RP}^2 \mathbin{\tilde{\times}} S^1$ is not diffeomorphic to $S^2 \times S^1$, because they have non-isomorphic fundamental groups. Moreover, $\mathbb{RP}^2 \mathbin{\tilde{\times}} S^1$ is not prime:

Exercise 10.3.37. The two manifolds $\mathbb{RP}^2 \mathbin{\tilde{\times}} S^1$ and $\mathbb{RP}^3 \# \mathbb{RP}^3$ are diffeomorphic.

We will soon see that $\mathbb{RP}^2 \mathbin{\tilde{\times}} S^1$ is the unique non-prime Seifert manifold. We now turn to the $\chi = 0$ case.

Proposition 10.3.38. *Every closed Seifert fibration with $\chi = 0$ and $e = 0$ is isomorphic to one of the seven listed in Table 10.3. These seven manifolds are all pairwise non-diffeomorphic, except*
$$\left(S^2, (2, 1), (2, 1), (2, -1), (2, -1)\right) \cong K \mathbin{\tilde{\times}} S^1.$$

M	$H_1(M, \mathbb{Z})$
$T \times S^1$	$\mathbb{Z}^3$
$K \; \tilde{\times} \; S^1$	$\mathbb{Z} \times \mathbb{Z}/_{2\mathbb{Z}} \times \mathbb{Z}/_{2\mathbb{Z}}$
$\left(S^2, (2,1), (2,1), (2,-1), (2,-1)\right)$	$\mathbb{Z} \times \mathbb{Z}/_{2\mathbb{Z}} \times \mathbb{Z}/_{2\mathbb{Z}}$
$\left(S^2, (3,1), (3,1), (3,-2)\right)$	$\mathbb{Z} \times \mathbb{Z}/_{3\mathbb{Z}}$
$\left(S^2, (2,1), (4,-1), (4,-1)\right)$	$\mathbb{Z} \times \mathbb{Z}/_{2\mathbb{Z}}$
$\left(S^2, (2,1), (3,1), (6,-5)\right)$	$\mathbb{Z}$
$\left(\mathbb{RP}^2, (2,1), (2,-1)\right)$	$\mathbb{Z}/_{4\mathbb{Z}} \times \mathbb{Z}/_{4\mathbb{Z}}$

Table 10.3. The seven closed Seifert fibrations with $\chi = 0$ and $e = 0$. Two of these manifolds are actually diffeomorphic, so we get six closed Seifert manifolds up to diffeomorphism, distinguished by their homology.

	$\chi > 0$	$\chi = 0$	$\chi < 0$
$e = 0$	2	6	∞
$e \neq 0$	∞	∞	∞
$\partial M \neq \varnothing$	1	2	∞

Table 10.4. The number of Seifert manifolds in each commensurability class.

Proof. The closed orbifolds with $\chi = 0$ are

T, K, $(\mathbb{RP}^2, 2, 2)$, $(S^2, 2, 2, 2, 2)$, $(S^2, 2, 3, 6)$, $(S^2, 3, 3, 3)$, $(S^2, 2, 4, 4)$.

It is easy to show that by imposing $e = 0$ we get the fibrations listed. The homology calculation is an easy exercise and luckily distinguishes all the manifolds except (of course) the two diffeomorphic ones (see Corollary 10.3.35). $\qquad\square$

Summing up, there are four commensurability classes containing finitely many manifolds, and their number is shown in Table 10.4.

10.3.12. Universal cover. We now determine the universal cover of all the closed Seifert manifolds. Nothing strange happens: we either get S^3, $S^2 \times \mathbb{R}$, or $\mathbb{R}^3$.

Proposition 10.3.39. *The universal cover of a closed Seifert manifold is shown in Table 10.5.*

	$\chi > 0$	$\chi = 0$	$\chi < 0$
$e = 0$	$S^2 \times \mathbb{R}$	$\mathbb{R}^3$	$\mathbb{R}^3$
$e \neq 0$	S^3	$\mathbb{R}^3$	$\mathbb{R}^3$

Table 10.5. The universal cover of a closed Seifert manifold depends on its invariants e and χ.

Proof. The universal cover of a circle bundle over a surface S with $\chi(S) \leqslant 0$ is a line bundle over the universal cover $\mathbb{R}^2$ of S. The line bundle is trivial since $\mathbb{R}^2$ is contractible and we get $\mathbb{R}^2 \times \mathbb{R} = \mathbb{R}^3$.   □

Corollary 10.3.40. *Every Seifert manifold M is irreducible or covered by $S^2 \times \mathbb{R}$. In the latter case M is diffeomorphic to $S^2 \times S^1$ or $\mathbb{RP}^2 \mathbin{\tilde{\times}} S^1$.*

Proof. If the universal cover is S^3 or $\mathbb{R}^3$, it is irreducible and hence also M is. When M has boundary, we apply the proof of Proposition 10.3.39 to the interior of M and find that its universal cover is $\mathbb{R}^3$.   □

Hence every Seifert manifold is prime except $\mathbb{RP}^2 \mathbin{\tilde{\times}} S^1 = \mathbb{RP}^3 \# \mathbb{RP}^3$.

10.3.13. Fibre-Parallel Dehn filling. We have investigated various topological properties of Seifert manifolds, and we are now curious: what happens if we perform a forbidden fibre-parallel Dehn filling? We start by looking at the basic block $P \times S^1$, where P is a pair-of-pants.

Proposition 10.3.41. *A fibre-parallel Dehn filling on $P \times S^1$ produces the connected sum of two solid tori; the fibres of $P \times S^1$ become the meridians of the solid tori.*

Proof. Let $\gamma \subset \partial P$ be the component whose torus $\gamma \times S^1$ is filled. Pick an essential arc α in P with both endpoints in γ: the fibered annulus $A = \alpha \times S^1$ closes up to a two-sphere S in the Dehn filling.
The sphere S separates the filled manifold into two portions, each diffeomorphic to $T \times [0, 1]$ with a two-handle attached along a non-trivial curve: this is a holed solid torus. Every fibre now bounds a disc there.   □

We now turn to arbitrary bundles.

Exercise 10.3.42. Let S be a compact surface with $b \geqslant 1$ boundary components. The fibre-parallel Dehn filling of $S \mathbin{\tilde{\times}} S^1$ is diffeomorphic to the connected sum of $b - 1$ solid tori and $-\chi(S) + 2 - b$ copies of $S^2 \times S^1$.

We now deduce a more general fact: a fibre-parallel Dehn filling of a Seifert manifold is a connected sum of lens spaces and solid tori.

Corollary 10.3.43. *The fiber-parallel Dehn filling of*

$$\bigl(S, (p_1, q_1), \ldots, (p_h, q_h)\bigr)$$

is diffeomorphic to the connected sum

$$L(p_1, q_1) \# \ldots \# L(p_h, q_h) \#_k (S^2 \times S^1) \#_{b-1} (D^2 \times S^1)$$

where S has b boundary components and $k = -\chi(S) + 2 - b$.

If S is orientable then k is twice the genus of S.

Corollary 10.3.44. *Every Dehn filling of a Seifert manifold is a connected sum of Seifert manifolds.*

If we have a Seifert manifold M with many boundary components and we Dehn fill some of them, we are guaranteed to produce a new Seifert manifold, unless one of the Dehn filling kills a fiber-parallel slope: in that case the manifold "degenerates" to a connected sum of lens spaces and solid tori. This is an interesting phenomenon to keep in mind, because it will reproduce in Chapter 15 in the hyperbolic world: we will show that by Dehn filling a cusped hyperbolic three-manifold we always get a new hyperbolic manifold if we avoid a finite number of "exceptional slopes" on every boundary torus; if some of the exceptional slopes is employed, the manifold is possibly not hyperbolic, and it typically "degenerates" and breaks into simpler pieces.

10.4. Classification

In the previous section we have classified the Seifert fibrations up to isomorphism and the Seifert manifolds up to commensurability. We now want to complete our study by classifying Seifert manifolds up to diffeomorphism: our final achievement will be Theorem 10.4.19 that determines precisely the Seifert manifolds that admit non-isomorphic fibrations.

The proof of Theorem 10.4.19 is not straightforward: we will apply different techniques to different classes of Seifert manifolds. We start by studying the fundamental group of Seifert manifolds, showing in particular that it fits into a nice short exact sequence. Then we study essential surfaces: we show that these can always be isotoped to be either in "vertical" or "horizontal" position with respect to the fibration.

In "most" cases, a Seifert manifold contains many incompressible vertical tori and these can be used to characterise the manifold. When the Seifert manifold is "small" it contains no such tori, and we must use different techniques: for instance we look at its fundamental group, which may be finite or infinite.

10.4.1. Fundamental group. We study the fundamental group of Seifert manifolds. Recall that every closed Seifert fibration has two fundamental invariants χ and e that determine the commensurability class of the Seifert manifold. Finite fundamental groups are easily detected:

Proposition 10.4.1. *A closed Seifert manifold has finite fundamental group $\iff$ it is covered by S^3 $\iff$ it has $\chi > 0$ and $e \neq 0$.*

Proof. A compact manifold has finite fundamental group $\iff$ its universal cover is also compact. Table 10.5 applies. □

Ordinary fibrations generate exact sequences in homotopy, and also Seifert fibrations do. Recall that a good orbifold has a well-defined fundamental group, and that all locally orientable surface orbifolds are good except (S^2, p) and (S^2, p_1, p_2) with $p_1 \neq p_2$. We will ignore bad orbifolds, since the Seifert manifolds fibering over them are lens spaces.

Proposition 10.4.2. *Let $M \to S$ be a Seifert fibration over a good orbifold S. There is an exact sequence*

$$1 \longrightarrow K \longrightarrow \pi_1(M) \longrightarrow \pi_1(S) \longrightarrow 1$$

where K is the normal cyclic subgroup of $\pi_1(M)$ generated by a regular fiber and $\pi_1(S)$ is the orbifold fundamental group of S.

Proof. The universal cover $\tilde{M}$ of M fibres over the universal cover $\tilde{S}$ of S. The fiber is a circle or a line, depending on whether $\pi_1(M)$ is finite or infinite, see Proposition 10.3.39.

The group $\pi_1(M)$ acts fibre-preservingly on $\tilde{M}$ and hence acts also on $\tilde{S}$ as a covering automorphism for $\tilde{S} \to S$. This induces a natural homomorphism $\pi_1(M) \to \pi_1(S)$. Its kernel consists of all deck transformations of $\tilde{M}$ that fix the fibres: these are precisely K. The homomorphism is surjective because every loop in S lifts to a loop in M. □

We now look more closely at the normal cyclic subgroup K.

Proposition 10.4.3. *The group K is infinite if and only if $\pi_1(M)$ is.*

Proof. The group K acts freely and proper discontinuously on each fiber of the universal cover $\tilde{M}$, and it quotients it to a circle. The fiber in $\tilde{M}$ is compact if and only if $\pi_1(M)$ is finite. □

Proposition 10.4.4. *The group K lies in the centre of $\pi_1(M)$ if and only if S is orientable or K has order two.*

Proof. If S is orientable, the manifold M is a Dehn filling of a product $S' \times S^1$ where the fiber is obviously in the centre, and it remains so after the Dehn filling. If S is non-orientable, pick an orientation-reversing loop α in S: there is a Klein bottle fibering above α and we get $\alpha g \alpha^{-1} = g^{-1}$ for the fiber $g \in K$. Therefore K is central if and only if $g^2 = e$. □

In many cases the subgroup K may be characterised intrinsically.

Proposition 10.4.5. *Let M be a Seifert manifold whose fundamental group is not virtually abelian. The subgroup $K \lhd \pi_1(M)$ is the unique maximal cyclic normal subgroup of $\pi_1(M)$.*

Proof. Suppose by contradiction that $\pi_1(M)$ contains a cyclic normal subgroup K' which is not contained in K. Therefore its image in $\pi_1(S)$ is a non-trivial cyclic normal subgroup.

If $\chi(S) < 0$ then $\mathrm{int}(S) = \mathbb{H}^2/_\Gamma$ for a discrete group $\Gamma = \pi_1(S)$ of isometries of $\mathbb{H}^2$. Corollary 5.1.10 implies that Γ does not contain non-trivial cyclic normal subgroups.

If $\chi(S) = 0$ then $\mathrm{int}(S) = \mathbb{R}^2/_\Gamma$ for a discrete group $\Gamma = \pi_1(S)$ of isometries of $\mathbb{R}^2$, which in fact may contain cyclic normal subgroups: so we need a more careful analysis. Since $\pi_1(M)$ is not virtually abelian, the surface S is closed and $e \neq 0$.

The image of K' in $\pi_1(S)$ is non-trivial, normal, and cyclic, and is hence infinite (finite cyclic groups are generated by rotations and cannot be normal). Therefore K' is infinite and it intersects non-trivially every finite-index subgroup of $\pi_1(M)$: up to substituting M with a finite cover we may suppose that S is a torus.

The image of K' is a non-trivial subgroup of $\pi_1(S) = \mathbb{Z} \times \mathbb{Z}$. Pick two elements $a' \in K', b \in \pi_1(M)$ whose images in $\pi_1(S)$ generate a finite-index subgroup of $\pi_1(S)$. Pick a generator $a \in K$. The three elements a, a', b generate a finite-index subgroup of $\pi_1(M)$. The element a is central by Proposition 10.4.4. We have $b^{-1}a'b = (a')^{\pm 1}$ because $K' = \mathbb{Z}$ is normal. The elements a, a', b^2 commute and generate a finite-index abelian $\mathbb{Z}^3 < \pi_1(M)$: a contradiction. $\square$

Corollary 10.4.6. *Let M be a Seifert manifold whose fundamental group is not virtually abelian. The centre of $\pi_1(M)$ is K if S is orientable and trivial otherwise.*

Proof. Use Propositions 10.4.4 and 10.4.3. $\square$

We have collected some important information on the fundamental group of Seifert manifolds. We now move from algebra to topology and study the essential surfaces in Seifert manifolds: we will prove a theorem analogous to Proposition 9.3.18, namely that every essential surface in a Seifert manifold is either horizontal or vertical.

10.4.2. Horizontal and vertical surfaces. We now want to study how Seifert manifolds may contain interesting surfaces. Let $M \to S$ be a Seifert fibration. A properly embedded surface $\Sigma \subset M$ is

- *vertical* if it is a union of some regular fibres,
- *horizontal* if it is transverse to all fibres.

If Σ is vertical, it is either an annulus, a torus, or a Klein bottle, projecting respectively to an arc, an orientation-preserving, or an orientation-reversing simple closed curve that avoids the cone points. Vertical surfaces are in 1-1 correspondence with 1-dimensional objects in S and are thus easily determined.
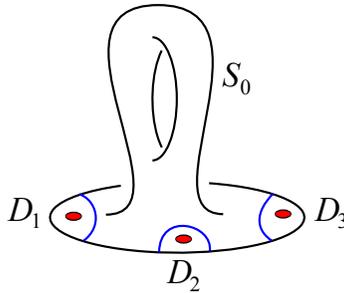
Figure 10.12. We cut S along k (blue) arcs to get $S = S_0 \cup D_1 \cup \ldots \cup D_k$ with D_i a disc containing the (red) cone point p_i: here $k = 3$.

Horizontal surfaces are more subtle. We first note that if Σ is horizontal, the natural projection $\Sigma \to S$ is an orbifold covering. When does M contain a horizontal surface? It certainly does when M has boundary.

Proposition 10.4.7. *Every Seifert fibration $M \to S$ with boundary contains a horizontal surface.*

Proof. We have $M = \big(S, (p_1, q_1), \ldots, (p_k, q_k)\big)$. We cut S along k arcs as in Figure 10.12 to get $S = S_0 \cup D_1 \cup \ldots \cup D_k$ with D_i a disc containing the cone point p_i. Let p be a common multiple of $p_1, \ldots, p_k$. The bundle over S_0 has no singular fibres and is hence $S_0 \mathbin{\widetilde{\times}} S^1$. The bundle $S_0 \mathbin{\widetilde{\times}} S^1$ contains a (possibly disconnected) horizontal surface Σ intersecting the fibres in any fixed number n of points (exercise) and we pick one Σ with $n = p$.

For every $i = 1, \ldots, k$ there is a standard fibered solid torus lying above D_i, attached to $S_0 \mathbin{\widetilde{\times}} S^1$ via a vertical annulus A. The meridian of this solid torus intersects A into p_i horizontal segments. Pick $\frac{p}{p_i}$ parallel meridians: both Σ and these meridians intersect A in p horizontal segments and can hence be glued to form a horizontal surface for M. $\qquad\square$

On closed Seifert manifolds the existence of a horizontal surface is fully detected by the Euler number.

Proposition 10.4.8. *A closed Seifert fibration $M \to S$ contains a horizontal surface if and only if $e = 0$.*

Proof. Suppose $M \to S$ contains a horizontal surface Σ. We pull-back the fibration along the orbifold covering $\Sigma \to S$ to get a $\tilde{M} \to \Sigma$. The horizontal surface Σ lifts to a section of $\tilde{M} \to \Sigma$. Since the fibration has a section, we get $e(\tilde{M}) = 0$ and hence $e(M) = 0$.

Conversely, suppose $e(M) = 0$. Drill one open fibered solid torus from M: we know by the previous proposition that the resulting fibration has

a horizontal surface Σ. Its boundary consists of parallel curves of some type (p, q) with $p > 0$. If we Dehn-fill the manifold by killing these curves we get another Seifert manifold M', to which the fibration and the section extend: therefore $e(M') = 0$. At most one pair (p, q) may produce a manifold with $e = 0$, and hence $M = M'$. $\qquad\square$

10.4.3. Essential surfaces. Recall the definition of essential surface from Sections 9.2.11 and 9.4.6. We now show that essential surfaces in irreducible Seifert manifolds are either vertical or horizontal.

Proposition 10.4.9. *Let $M \to S$ be a Seifert fibration and M be irreducible. Every essential surface Σ is isotopic to a vertical or horizontal surface.*

Proof. See M as a Dehn filling of $S \mathbin{\overset{(\sim)}{\times}} S^1$. We suppose that Σ intersects transversely the cores of the filling solid tori in the minimum number of points up to isotopy. After an isotopy Σ intersects the filling solid tori into parallel horizontal discs.

Decompose S into one 0-handle and some g one-handles. We see the 0-handle as a $2g$-gon and each 1-handle as a rectangle. Above each edge of the $2g$-gon or of a rectangle there is a vertical annulus A. We put Σ in transverse position with respect to these vertical annuli and up to isotopy we may suppose that $\Sigma \cap A$ consists either of vertical fibres or horizontal arcs for each A. Indeed we can easily eliminate trivial circles (because Σ is incompressible and M is irreducible) and arcs forming bigons with ∂A (because Σ is ∂-irreducible and has minimal intersection with the solid tori cores).

Above every polygonal handle of S there is a prismatic solid torus W. The closed curves $\Sigma \cap \partial W$ are made of horizontal or vertical arcs and are hence essential, forming some parallel slopes in ∂W. We can suppose that $\Sigma \cap W$ consists of essential discs or incompressible surfaces (if there is a compressing disc D inside W, then $\partial D \subset \Sigma$ bounds a disc in Σ which we isotope to D reducing the intersection with the vertical annuli). We can also suppose that $\Sigma \cap W$ is ∂-incompressible with respect to every vertical annulus $A \subset \partial W$, in the sense that there is no ∂-compressing disc D with $\partial D \subset A \cup \Sigma$ (otherwise we isotope Σ and reduce intersections).

By Proposition 9.3.16 the surface $\Sigma \cap W$ consists either of essential discs or ∂-parallel annuli (not both). In the first case $\Sigma \cap \partial W$ consists of horizontal curves and $\Sigma \cap W$ consists of horizontal discs. In the second case, since the annuli are ∂-incompressible with respect to the vertical annuli in ∂W, their boundaries must be vertical circles lying in distinct vertical annuli. Up to isotopy $\Sigma \cap W$ consists of vertical annuli.

We have decomposed Σ into horizontal discs or vertical annuli, and both cannot coexist. Hence Σ is either horizontal or vertical. $\qquad\square$

Corollary 10.4.10. *Every Seifert manifold is irreducible and ∂-irreducible, except $S^2 \times S^1$, $\mathbb{RP}^2 \mathbin{\widetilde{\times}} S^1$, and $D \times S^1$.*
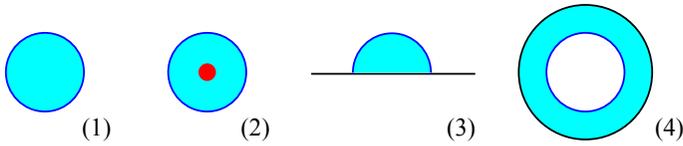
Figure 10.13. A vertical torus or annulus is essential, unless its projection is: the boundary of a disc containing zero (1) or one (2) singular cone points, an arc parallel to the boundary (3), or a ∂-parallel closed curve (4).
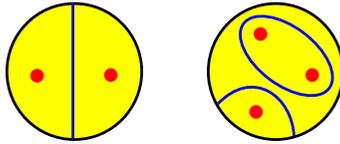

Figure 10.14. Essential vertical annuli and tori in a Seifert manifold fibering over a disc with 2 or 3 singular points.

Proof. We already know it is irreducible with the first two exceptions. If it contains an essential disc D, it contains a horizontal one which covers the base surface S of the fibration, hence $\chi(S) > 0$ and we get $D \times S^1$. □

We now prove a converse to Proposition 10.4.9.

Proposition 10.4.11. *Let $M \to S$ be a Seifert fibration and M be irreducible. Let $\Sigma \subset M$ be an orientable connected surface. Suppose that*

- *Σ is horizontal, or*
- *Σ is vertical and its projection is not as in Figure 10.13.*

Then Σ is essential.

Proof. If Σ is horizontal, it finitely covers S and hence $\pi_1(\Sigma)$ injects in $\pi_1(S)$. Therefore it injects also in $\pi_1(M)$ and Σ is incompressible. By doubling everything along ∂M we get a horizontal $D\Sigma$ inside DM, which must also be incompressible: this implies that Σ is ∂-incompressible. It is also clearly not ∂-parallel, so it is essential.

If Σ is vertical, by cutting along it we get one or two Seifert fibrations. Since Σ is not as in Figure 10.13, the base orbifolds of these fibrations have $\chi \leqslant 0$, so their boundary is incompressible by Corollary 10.4.10. Therefore Σ is incompressible and ∂-incompressible. Moreover it is not a ∂-parallel torus because none of the cut Seifert manifold is $A \times S^1$. □

Example 10.4.12. The curves in Figure 10.14 determine vertical essential annuli and tori in Seifert manifolds fibering over the disc with 2 or 3 singular fibres.

We can now fully detect essential tori in almost all Seifert manifolds.

Corollary 10.4.13. *Let $M \to S$ be a Seifert fibration and M be irre-ducible. If M is not covered by a 3-torus or $T \times [0, 1]$, every essential torus or annulus is vertical.*

Proof. Horizontal tori or annuli may arise only when $\chi = 0$ and $\partial M \neq 0$ or $e = 0$. □

10.4.4. Simple Seifert manifolds. We want to classify Seifert mani-folds up to diffeomorphism. To do so, we group them into some classes, and use different techniques on each class. Recall that a 3-manifold is *simple* if it contains no essential sphere, disc, annulus, or torus.

Proposition 10.4.14. *A Seifert manifold is simple $\iff$ it fibres over S^2 with at most three singular fibres and is not covered by $S^2 \times S^1$ or the 3-torus.*

Proof. Let $M \to S$ be a Seifert fibration. The manifold M contains an essential vertical torus or annulus $\iff$ the orbifold S contains an arc or an orientation-preserving simple closed curve that is not as in Figure 10.13. The only orbifolds that do not contain such arcs or curves are: D with at most one fiber, S^2 with at most three fibres, and $\mathbb{RP}^2$ with at most one fiber. In the first case M is a solid torus, which contains an essential disc, and in the third case M also fibres over S^2 with at most three singular fibres, see Corollary 10.3.35.

We are left to consider horizontal closed essential surfaces with $\chi \geqslant 0$. These arise only when $e = 0$ and $\chi(S) \geqslant 0$, *i.e.* when M is covered by $S^2 \times S^1$ or the 3-torus. □

We note in particular that all the simple Seifert manifolds are closed.

Remark 10.4.15. Among the nine commensurability classes of Seifert manifolds, three contain simple manifolds: those with empty boundary and $e \neq 0$. The Euler characteristic χ of a sphere with 3 singular fibres may in fact be positive, null, or negative. In particular there are simple Seifert manifolds with finite and with infinite fundamental group.

10.4.5. Seifert manifolds with boundary. To classify Seifert mani-folds up to diffeomorphism, we will show that (except a few explicit excep-tions) a generic Seifert manifold has a unique fibration up to isomorphism. We start with the easier non-empty-boundary case.

Proposition 10.4.16. *Every Seifert manifold M with non-empty bound-ary admits only one fibration up to isomorphism, except in the following cases:*

- $M = D \times S^1$ *fibres as* $\big(D, (p, q)\big)$,
- $M = S \mathbin{\widetilde{\times}} S^1$ *fibres as* $\big(D, (2, 1), (2, -1)\big)$.

Here S is the Möbius strip.

Proof. If M fibres over an orbifold S with $\chi(S) \geqslant 0$, we have already proved this in Section 10.3.11. Suppose that M has two fibrations $M \to S$ and $M \to S'$ with $\chi(S), \chi(S') < 0$.

Pick a minimal collection of properly embedded arcs in S that avoid the cone points and decompose S into discs, each containing at most one cone point. Each arc determines an essential vertical annulus in $M \to S$. The complement of these annuli consists of vertical fibered solid tori.

By Proposition 10.4.11 these annuli are essential and by Proposition 10.4.9 they are isotopic to vertical annuli with respect to the other fibration $M \to S'$ (they cannot be horizontal since $\chi(S') < 0$). Therefore we may isotope the fibration $M \to S'$ so that the annuli are vertical in both fibrations, and by further isotoping we can in fact suppose that the two fibrations coincide on a neighbourhood of these annuli.

The complement of this neighbourhood consists of solid tori. The two fibrations are obtained from the same fibration by Dehn-filling along the same slopes, and hence they are isomorphic. $\qquad\square$

10.4.6. Seifert manifolds with infinite fundamental group. We now turn to closed Seifert manifolds. We start by examining the $\chi \leqslant 0$ case.

Proposition 10.4.17. *Every closed Seifert manifold M not covered by S^3 or $S^2 \times S^1$ admits only one fibration up to isomorphism, except:*

$$\left(S^2, (2, 1), (2, 1), (2, -1), (2, -1)\right) \cong K \;\widetilde{\times}\; S^1.$$

Proof. If the Seifert manifold M is covered by the 3-torus, we have already proved this in Proposition 10.3.38. Suppose that M is not covered by S^3, $S^2 \times S^1$, or the 3-torus. Let M have two fibrations $M \to S$ and $M \to S'$.

We first suppose that M contains an essential torus. We try to proceed as in the proof of Proposition 10.4.16 using vertical tori instead of annuli. Essential tori are vertical to both fibrations by Corollary 10.4.13. In particular both S and S' are not spheres with at most three singular points.

Let $\dot{S} \subset S$ be S without its singular points. Pick two multicurves $C_1, C_2 \subset \dot{S}$ without puncture-parallel components, that fill $\dot{S}$ and intersect minimally. By filling $\dot{S}$ we mean that $\dot{S} \setminus (C_1 \cup C_2)$ consists of discs or once-punctured discs, and minimality implies that no such disc is an unpunctured bigon.

Let T_1, T_2 be the collections of tori fibering above C_1, C_2. The set T_1 consists of disjoint essential tori, and we can isotope the fibration $M \to S'$ so that they are vertical with respect to the fibrations $M \to S$, $M \to S'$.

We turn to the tori T_2, that are vertical with respect to $M \to S$. Up to isotopy, each simple closed curve in $T_1 \cap T_2$ is either horizontal or vertical with respect to $M \to S'$. Each annulus in $T_2 \setminus T_1$ is essential in $M \setminus T_1$ and hence it is correspondingly horizontal or vertical after an isotopy. Horizontal annuli would glue to a horizontal torus in T_2, which is

excluded: hence the annuli are vertical with respect to $M \to S'$. We can therefore isotope the fibration $M \to S'$ so that the tori T_2 are also vertical.

Now $T_1 \cup T_2$ is vertical with respect to both fibrations. Since C_1 and C_2 fill, every torus in T_1 intersects one in T_2 and viceversa, so $T_1 \cup T_2$ is made of vertical annuli attached along the vertical curves $T_1 \cap T_2$. Up to isotopy the fibrations $M \to S$ and $M \to S'$ coincide on the vertical annuli, hence on $T_1 \cup T_2$, hence on a neighbourhood of $T_1 \cup T_2$.

The complement $M \setminus (T_1 \cup T_2)$ consists of vertical solid tori and we conclude as in Proposition 10.4.16 that the two fibrations are isomorphic.

We are left to consider the case where M contains no essential tori, and hence is a simple manifold. Both S, S' are spheres with exactly three singular points by Proposition 10.4.14 (not less than three singular points since M is not covered by S^3 or $S^2 \times S^1$).

By Corollary 10.4.6, the fundamental groups $\pi_1(S)$ and $\pi_1(S')$ are both isomorphic to $\pi_1(M)$ quotiented by its centre. Exercise 3.6.13 implies that the two orbifolds are isomorphic, so $S = S' = (S^2, p_1, p_2, p_3)$ for some $p_1, p_2, p_3 \geqslant 2$. Consider the exact sequence

$$1 \longrightarrow K \longrightarrow \pi_1(M) \longrightarrow \pi_1(S) \longrightarrow 1.$$

Fix a generator $l \in K$ and coherently an orientation for the fibres of both fibrations $M \to S$ and $M \to S'$. The Von Dyck group $\pi_1(S^2, p_1, p_2, p_3)$ has a presentation

$$\langle\, r_1, r_2, r_3 \mid r_1^{p_1}, r_2^{p_2}, r_3^{p_3}, r_1 r_2 r_3 \,\rangle$$

and the three generators r_1, r_2, r_3 are intrinsically determined up to simultaneous conjugation or inversion, see Exercise 3.6.13. Fix three lifts m_1, m_2, m_3 of r_1, r_2, r_3 in $\pi_1(M)$ with $m_1 m_2 m_3 = 1$: these lifts determine sections for the fibrations $M \to \dot{S}$ and $M \to \dot{S}'$ with boundary meridians m_1, m_2, m_3. We use these sections to determine the parameters q_i and q_i' in both fibrations, for $i = 1, 2, 3$.

The centre K is infinite since $\pi_1(M)$ is, see Proposition 10.4.3. Therefore $m_i^{p_i} \in K$ equals $l^{q_i''}$ for some unique $q_i'' \in \mathbb{Z}$. By construction we have $q_i'' = q_i$ and $q_i'' = q_i'$, hence $q_i = q_i'$. The two fibrations $M \to S$ and $M \to S'$ are isomorphic. $\square$

10.4.7. Seifert manifolds with finite fundamental group. This long journey through Seifert manifolds is almost finished: it remains to classify the Seifert manifolds with finite fundamental group, that is that are covered by the three-sphere.

Proposition 10.4.18. *Let M be a Seifert manifold covered by S^3 that is not a lens space. It has a unique fibration over one of the orbifolds*

$$S = (S^2, 2, 2, p), \quad (S^2, 2, 3, 3), \quad (S^2, 2, 3, 4), \quad (S^2, 2, 3, 5)$$

for some $p \geqslant 2$. The centre of M is the cyclic group K in

$$1 \longrightarrow K \longrightarrow \pi_1(M) \longrightarrow \pi_1(S) \longrightarrow 1.$$

fibration	$e(M)$	$|\pi_1(S)|$	$|\pi_1(M)|$	$|H_1(M, \mathbb{Z})|$
$\left(S^2, (2, 1), (2, 1), (p, q)\right)$	$\frac{p+q}{p}$	$2p$	$4p|p + q|$	$4|p + q|$
$\left(S^2, (2, 1), (3, 1), (3, q)\right)$	$\frac{5+2q}{6}$	12	$24|5 + 2q|$	$3|5 + 2q|$
$\left(S^2, (2, 1), (3, 1), (4, q)\right)$	$\frac{10+3q}{12}$	24	$48|10 + 3q|$	$2|10 + 3q|$
$\left(S^2, (2, 1), (3, 1), (5, q)\right)$	$\frac{25+6q}{30}$	60	$120|25 + 6q|$	$|25 + 6q|$

Table 10.6. The non-lens Seifert manifolds M with finite fundamental group. For each fixed base orbifold S, the fibration is determined by the order of $H_1(M, \mathbb{Z})$: when two different parameters q give the same order of $H_1(M, \mathbb{Z})$, the fibrations are actually the same (with opposite orientations and hence Euler numbers).

The manifold M is determined by S and $H_1(M)$, see Table 10.6.

Proof. We have $M \to S$ with $\chi(S) > 0$ and hence S is either a sphere with at most two singular points (so M is a lens space), one of the orbifolds listed, or $\mathbb{RP}^2$ with one singular point; in the latter case M also fibres over $(S^2, 2, 2, p)$ by Corollary 10.3.35.

The subgroup K is central by Proposition 10.4.4. Suppose that $\pi_1(M)$ contains a central element disjoint from K: its image in $\pi_1(S)$ is a non-trivial central element. However S is spherical and $\pi_1(S)$ is a non-cyclic group of rotations of S^2: two rotations with different axis never commute.

The subgroup $K < \pi_1(M)$ is intrinsically determined as the centre, hence the quotient $\pi_1(S)$ is also determined, and S also is by Exercise 3.6.13. Therefore M cannot fiber on two different orbifolds in the list.

To conclude, we prove that the fibration $M \to S$ is determined by S and the finite number $|H_1(M, \mathbb{Z})|$. We analyse each case separately. First, we transform the fibration as in Table 10.6: using the move from Proposition 10.3.11-(8) we manage to transform each $(2, q')$ in $(2, 1)$ and $(3, q')$ in either $(3, 1)$ or $(3, 2)$, and by reversing orientation (see Remark 10.3.14) we transform $(3, 2)$ into $(3, -2)$ and hence into $(3, 1)$.

Using Proposition 10.3.15 we see that $|H_1(M)|$ is as in Table 10.6. The only cases where different values of q give the same $|H_1(M)|$ are the following:

- $S = (S^2, 2, 2, p)$ and $|p + q| = |p + q'|$, hence $q' = -q - 2p$,
- $S = (S^2, 2, 3, 3)$ and $|5 + 2q| = |5 + 2q'|$, hence $q' = -q - 5$ which implies $q, q' \equiv 2 \pmod 3$.

In these cases it is easy to verify that the fibrations with q and q' are isomorphic using Proposition 10.3.11-(8) and Remark 10.3.14.

The order $|\pi_1(M)|$ in Table 10.6 is obtained using the formula

$$|\pi_1(M)| = \frac{4e(M)}{\chi(S)^2}$$

that we now prove. Since the orbifolds are good, the fibration is covered horizontally by a fibration over S^2, which is in turn covered vertically by $\left(S^2, (1,1)\right) = S^3$. The total degree is $d = d_{\mathrm{h}} \cdot d_{\mathrm{v}}$. Proposition 10.3.22 gives

$$d_{\mathrm{h}} = \frac{\chi(S^2)}{\chi(S)} = \frac{2}{\chi(S)}, \quad 1 = e(S^2, (1,1)) = \frac{d_{\mathrm{h}}}{d_{\mathrm{v}}} \cdot e(M).$$

Therefore $d_{\mathrm{v}} = d_{\mathrm{h}} \cdot e(M)$ and

$$d = d_{\mathrm{h}} \cdot d_{\mathrm{v}} = d_{\mathrm{h}}^2 \cdot e(M) = \frac{4e(M)}{\chi(S)^2}$$

as required. $\square$

10.4.8. Summary. We now summarise the topological classification of Seifert manifolds in a single statement.

The Seifert fibrations $M \to S$ are fully classified by Proposition 10.3.11 and Corollary 10.3.13. The latter says that two Seifert fibrations

$$\left(S, (p_1, q_1), \dots, (p_h, q_h)\right), \quad \left(S', (p_1', q_1'), \dots, (p_{h'}', q_{h'}')\right)$$

with $p_i, p_i' \geqslant 2$ are orientation-preservingly isomorphic if and only if $S = S'$, $h = h'$, $e = e'$, and up to reordering $p_i = p_i'$ and $q_i \equiv q_i' \pmod{p_i}$ for all i.

We easily understand when two Seifert fibrations are isomorphic. To classify Seifert manifolds up to diffeomorphism it only remains to understand which Seifert manifolds can have non-isomorphic fibrations. A long discussion has shown the following. We write $S^2 \times S^1$ as the lens space $L(0, 1)$.

Theorem 10.4.19. *Every Seifert manifold has a unique Seifert fibration up to isomorphism, except the following:*

- *$L(p, q)$ fibres over S^2 with $\leqslant 2$ singular points in many ways,*
- *$D \times S^1$ fibres over D with $\leqslant 1$ singular point in many ways,*
- *$\left(D, (2,1), (2,-1)\right) \cong S \mathbin{\widetilde{\times}} S^1$,*
- *$\left(S^2, (2,1), (2,-1), (p,q)\right) \cong \left(\mathbb{RP}^2, (q,p)\right)$,*
- *$\left(S^2, (2,1), (2,1), (2,-1), (2,-1)\right) \cong K \mathbin{\widetilde{\times}} S^1$.*

Here S and K are the Möbius strip and the Klein bottle.

10.4.9. References. The main sources that we have consulted for this long chapter are Hatcher [26] and Scott [52]. Some material can also be found in Fomenko – Matveev [18]. Two classical references are Seifert's original paper [53] and Orlik [43].

Constructions of three-manifolds

In the previous chapter we have introduced and fully classified an important family of three-manifolds called *Seifert manifolds*, and we now address the following question: how can we construct more three-manifolds?

The most popular techniques employed for the construction and manipulation of three-manifolds are of cut-and-paste type: we build three-manifolds by gluing some blocks altogether, and we try to describe both the blocks and their gluing with some reasonable combinatorial formalism.

The choice of the right blocks is of course fundamental, and different choices lead to quite different environments. The first reasonable option may be to use tetrahedra as blocks, and in that case we talk about *triangulations* of three-manifolds: this construction has a strong combinatorial flavour and can be easily carried on by a computer. Other choices involve blocks without "ridges", that is manifolds with boundary: by using handlebodies we get *Heegaard splittings*, with knot/link complements and solid tori we get *Dehn surgery*, and with product manifolds $\Sigma \times [-1, 1]$ we get *surface bundles*. We introduce here all these topological constructions.

We end this chapter by showing that every prime three-manifold has a canonical decomposition along disjoint embedded tori, called the *geometric decomposition* – the reason for adopting this name will be evident in the next chapters.

11.1. Heegaard splittings

A *Heegard splitting* is a decomposition of a closed three-manifold in two manifolds of the simple kinds, the handlebodies.

11.1.1. Definition. The following proposition is quite surprising, because it shows that every closed orientable three-manifold decomposes into two pieces of a very simple type.

Proposition 11.1.1. *Every closed orientable 3-manifold M decomposes into two handlebodies of some genus g.*

Proof. The 3-manifold M has a handle decomposition with 0-, 1-, 2-, and 3-handles. The 0- and 1-handles altogether form a handlebody. The 2- and 3-handles may be turned upside down to form a handle decomposition into 0- and 1-handles, so another handlebody. □
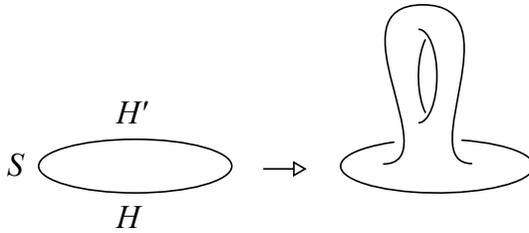
Figure 11.1. A stabilisation of a Heegaard splitting.

A decomposition of M into two handlebodies is called a *Heegaard splitting*. The two handlebodies have necessarily the same genus g, since their boundaries are glued together and are hence diffeomorphic surfaces.

Definition 11.1.2. The *Heegaard genus* $g(M)$ of a closed orientable M is the minimum genus of a Heegaard splitting for M.

11.1.2. Examples. The manifolds of genus zero and one are perfectly understood.

Proposition 11.1.3. *The 3-sphere has genus zero, lens spaces (except S^3) have genus one, and all the other closed orientable three-manifolds have genus at least two.*

Proof. By gluing two discs we get a 3-sphere, and by definition by gluing two solid tori we get a lens space. □

Classifying closed manifolds of genus two is a much harder task. We limit ourselves to a class of examples.

Exercise 11.1.4. Every Seifert manifold $\big(S^2, (p_1, q_1), (p_2, q_2), (p_3, q_3)\big)$ with $p_1, p_2, p_3 \geqslant 2$ has genus two.

Hint. Pick two singular fibres and connect them with a horizontal arc. A regular neighbourhood of this graph is a genus-two handlebody: prove that its complement is also a handlebody. □

11.1.3. Stabilisation. The same manifold may have various non isotopic Heegaard splittings: for instance there is a simple move that modifies a Heegaard splitting by increasing its genus in a somehow trivial way.

A *stabilisation* is a move as in Figure 11.1 that transforms a Heegaard splitting $M = H \cup_S H'$ of genus g into one of genus $g + 1$ of the same manifold M. We add an unknotted handle to the surface $S = \partial H = \partial H'$, so that both H and H' are transformed into handlebodies of genus $g + 1$.

Example 11.1.5. The complement of a standardly embedded genus-g handlebody in S^3 as in Figure 9.6 is another handlebody, and together they form a genus-g Heegaard splitting of S^3 obtained by stabilising g times the genus-0 Heegaard splitting of S^3.

11.1.4. Triangulations. Smooth triangulations are somehow related to Heegaard splittings, at least in one direction.

Every compact manifold admits a smooth triangulation (see Section 1.7.7) and in dimension three we can luckily prove a converse statement, much as we did in dimension two (in Section 6.1.5).

Let X be a three-dimensional pure simplicial complex, where: every face is incident to two tetrahedra, every edge is contained in a cycle of adjacent tetrahedra, and the link of every vertex is a sphere.

Proposition 11.1.6. *The complex X is the smooth triangulation of a closed three-manifold M, unique up to diffeomorphism.*

Proof. By dualising X we get a handle decomposition: tetrahedra, triangles, edges, and vertices determine 0-, 1-, 2-, and 3-handles. This procedure constructs a smooth closed three-manifold M triangulated by X. The way handles are attached is determined up to isotopy, and hence M is determined up to diffeomorphism.   □

As we did with surfaces, it is worth noting that this procedure (getting a unique smooth structure from a simplicial complex) does not work in all dimensions (here we used implicitly the non-obvious Proposition 9.2.1).

Topologists usually prefer to loosen the notion of "triangulation" by allowing self- and multiple adjacencies between tetrahedra, see Section 14.1.2. What is important to note here is that, no matter what the definition of "triangulation" is, a triangulated three-manifold always has a well-defined smooth structure determined only by the combinatorics of the triangulation. This shows that smooth three-manifolds can be treated combinatorially, for instance by a computer.

A relation between triangulations and Heegaard splittings is the following: a smooth triangulation of a closed manifold with t tetrahedra gives rise to a dual handle decomposition and hence to a Heegaard splitting of genus t. A fundamental difference between the two notions is that there is a bounded number of closed three-manifolds triangulated by at most t tetrahedra for every t, while there are infinitely many manifolds with Heegaard genus at most g for every $g \geqslant 1$.

11.2. Knots and links

Knots and links are fundamental and beautiful objects in geometric topology. Knots and links in S^3 have a combinatorial and mildly two-dimensional nature because they can be treated as planar diagrams, but they should be considered as intrinsically three-dimensional objects.

11.2.1. Definition. A *link* in a 3-manifold M is a compact submanifold of dimension one. Being compact, it consists of finitely many circles, and a connected link is called a *knot*. Links and knots are usually considered up to ambient isotopy.
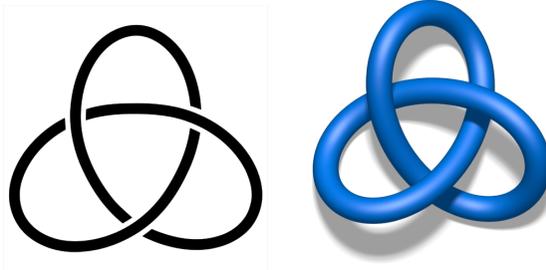
Figure 11.2. Every knot may be described via a planar dia-
gram with over/under crossings (left) and its tubular neigh-
bourhood is a knotted solid torus (right).

Every link in S^3 can be projected to a plane: if the link is generic
with respect to the projection, its image is a *diagram* as in Figure 11.2-
(left). Every link in S^3 can be described via a diagram and can be thus
treated combinatorially, but it is important to note that the same link can
be represented via infinitely many diagrams and it is hard in general to tell
whether two given diagrams represent the same link. The *trivial knot* (or
unknot) $K \subset S^3$ is the one that has a planar diagram without crossings.

11.2.2. Chirality. The *mirror image* of a knot or link is obtained by
reflecting it with respect to any plane in S^3. On a diagram, this operation
is realised simply by inverting all crossings simultaneously.

A knot is *chiral* if it is not isotopic to its mirror image, and *achiral*
otherwise. For instance, the trivial knot is achiral, but the trefoil knot
(shown in Figure 11.2) is chiral. As in the rest of this book, we mostly
ignore orientation issues and often consider implicitly two mirrored knots
or links as equivalent.

11.2.3. Link complement. The tubular neighbourhood of a link L in
an orientable M consists of solid tori by Proposition 1.1.13, and the *link
complement* of L is the three-manifold with boundary obtained from M
by removing the interiors of these solid tori. If M is compact, the link
complement also is.

Proposition 11.2.1. *Let $K \subset S^3$ be a knot. The complement of K is
an irreducible manifold. Moreover, the following facts are equivalent:*

(1) *K is trivial,*
(2) *K bounds a disc in S^3,*
(3) *the complement of K is a solid torus.*

Proof. The complement of K is irreducible by Proposition 9.2.15.
The implication (1)$\Rightarrow$(2) is the smooth Jordan curve theorem, and (2)$\Rightarrow$(1)
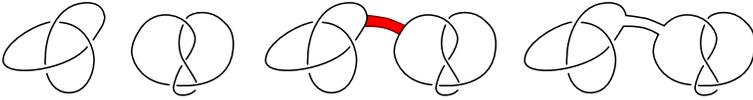holds because all discs in a connected three-manifold are isotopic.

Figure 11.3. Connected sum of knots: put two knots in disjoint balls and connect them with a band as shown.

A trivial knot thickens to a standardly embedded solid torus, whose complement is a solid torus: hence (1)⇒(3). Conversely, if the complement of K is a solid torus then S^3 decomposes into two solid tori as $S^3 = L(1, q)$. The meridian of the complement solid torus extends to a disc with boundary K, hence (3)⇒(2). □

Corollary 11.2.2. *All Heegaard tori for S^3 are isotopic*

Proof. A Heegaard torus $T \subset S^3$ is by definition a torus that decomposes S^3 into two solid tori. By the proposition the core of one of these solid tori is always isotopic to a trivial knot and hence T is isotopic to a standard torus. □

A link complement in S^3 can be reducible: this holds precisely when there is a sphere $\Sigma \subset S^3$ disjoint from the link L that cut S^3 into two balls, each containing some components of L. In that case we say that the link is *split*.

Corollary 11.2.3. *Let $L \subset S^3$ be a non-trivial knot or non-split link. The link complement is Haken.*

Proof. The link complement M is irreducible by assumption; by Proposition 9.2.30, if M contains an essential disc then it is a solid torus and hence L is the trivial knot, which is excluded. Now M is Haken by Corollary 9.4.5. □

11.2.4. Prime knots. There is an operation on knots in S^3 called *connected sum*, similar to the one on manifolds and described in Figure 11.3. A connected sum is *trivial* if one of the two knots is trivial; a non-trivial knot is *prime* if it is not the result of a non-trivial connected sum, and *composite* otherwise.

The *crossing number* of a knot is the minimum number of crossings in a diagram describing it. Prime knots with small crossing number have been tabulated since the XIX century: at present all the prime knots with up to 16 crossing have been classified (by Hoste – Thistlethwaite – Weeks [**29**] in 1998) and the first 14 numbers of them are listed in Table 11.1. The prime knots with $\leqslant 7$ crossings are shown in Figure 11.4: the first three in the list are the unknot, the *trefoil*, and the *figure-eight knot*.

c	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
n	1	0	0	1	1	2	3	7	21	49	165	552	2176	9988	46972

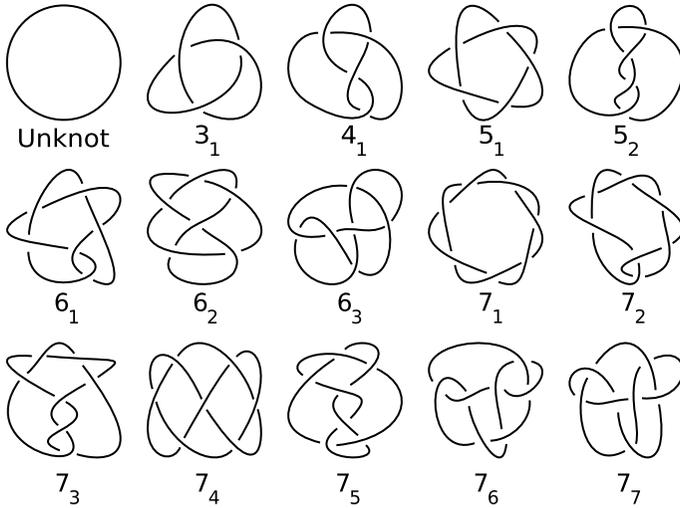Table 11.1. The number n of prime knots with c crossings, for all $c \leqslant 14$.

Figure 11.4. Prime knots in S^3 with at most 7 crossings. The knots 3_1 and 4_1 are the *trefoil* and the *figure-eight* knots respectively.
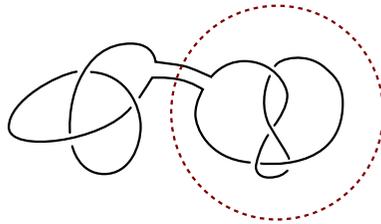

Figure 11.5. A connected sum produces a sphere (dotted in the figure) intersecting the knot in two points.

A connected sum gives rise to a sphere which intersects the new knot in two points, see Figure 11.5. The sphere intersects the knot complement into an essential annulus.
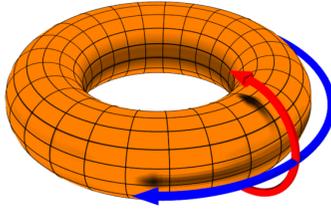
Figure 11.6. The meridian m (red) and longitude l (blue) of a standard torus are oriented as shown here, like a right-hand screw.
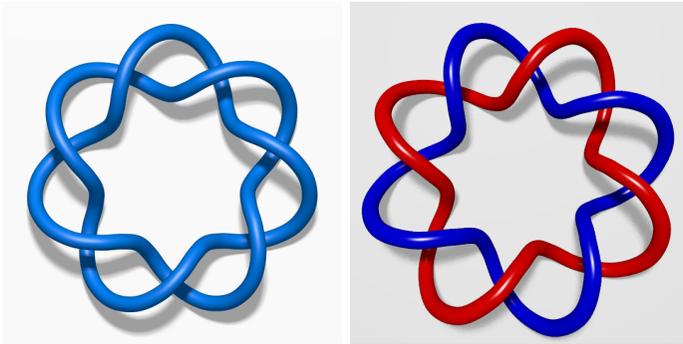
Figure 11.7. The torus knot $(2, -7)$ and the torus link $(2, -8)$.

11.2.5. Torus knots. We now introduce a simple and natural family of prime knots. Consider the standardly embedded torus $T \subset S^3$ with meridian m and longitude l oriented as in Figure 11.6. A pair (p, q) of coprime integers determines a simple closed curve in T that is homologically $qm + pl$, and such a curve is called a (p, q)-*torus knot*. More generally, a pair of integers (a, b) determines a multicurve in T that is homologically $bm + al$, called an (a, b)-*torus link*.

For instance, the trefoil knot is a $(2, -3)$-torus knot. More examples are in Figure 11.7. The knots 3_1, 5_1, 7_1 from Figure 11.4 are torus knots.

Exercise 11.2.4. If $|p| \leqslant 1$ or $|q| \leqslant 1$ the torus knot is the unknot. The torus knots with parameters (p, q), $(-p, -q)$, and (q, p) are isotopic, the torus knot with parameters $(p, -q)$ is the mirror image of (p, q).

The complement of a torus knot is a quite simple kind of Seifert manifold. In the following we suppose that $p, q \geqslant 1$ are coprime.
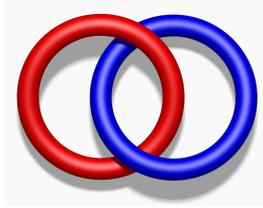
Figure 11.8. The Hopf link is the $(2, 2)$-torus link.

Proposition 11.2.5. *The complement of a (p, q)-torus knot is a Seifert manifold fibering over the orbifold (D, p, q). More precisely, it is*

$$\big(D, (p, r), (q, s)\big)$$

where (r, s) is any pair such that $ps + qr = 1$.

Proof. Let $K \subset T \subset S^3$ be the (p, q)-torus knot. The pair $(r, -s)$ determines another simple closed curve $\alpha \subset T$ that intersects K in one point. The complement of K in a tubular neighbourhood $T \times [-1, 1]$ of T is diffeomorphic to $P \times S^1$, where P is a pair of pants. On the tori $T \times \{-1\}$ and $T \times \{1\}$ the curves $(\partial P \times \{\text{pt}\}) \cap (T \times \{-1\})$ and $\{\text{pt}\} \times S^1$ are isotopic to α and K.

The complement of $T \times [-1, 1]$ in S^3 consists of two solid tori, with meridians $(1, 0)$ and $(0, 1)$. Read in the basis (α, K) the meridians are (q, s) and (p, r). The complement of K in S^3 is obtained from $P \times S^1$ by filling these curves and hence we get $\big(D, (p, r), (q, s)\big)$.   □

In particular the complement of the trefoil knot is $\big(D, (2, 1), (3, 1)\big)$. If we suppose $p, q \geqslant 2$, two torus knots with distinct (unordered) parameters have non-diffeomorphic complements and hence are not isotopic. Our intuition says that a torus knot is prime, and we can prove this rigorously.

Proposition 11.2.6. *Every torus knot is prime.*

Proof. If K is obtained as a non-trivial connected sum, there is an essential annulus in the complement, whose boundary curves are meridians of K. There is only one essential annulus in $\big(D, (p, r), (q, s)\big)$ and its boundary curves are not meridians.   □

The $(2, 2)$-torus link is called the *Hopf link* and is drawn in Figure 11.8.

Exercise 11.2.7. The complement of the Hopf link is $T \times [0, 1]$. Any two distinct fibres of the Hopf fibration form a Hopf link.
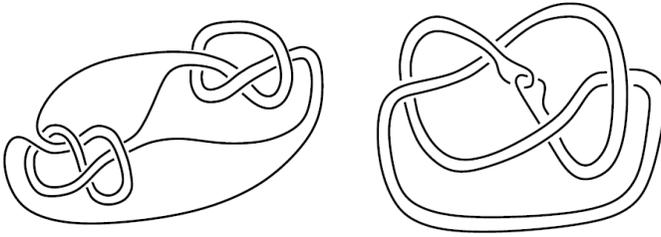
Figure 11.9. Two satellite knots.

11.2.6. Satellite knots. We have discovered that che composite knots contain essential annuli in their complement, and we now look for essential tori. It is convenient to introduce the following definition.

Definition 11.2.8. A knot $K \subset S^3$ is a *satellite* if its complement contains an essential torus.

We now formulate an equivalent and more inspiring definition of satellite knots which justifies the terminology. A knot in a solid torus $D^2 \times S^1$ is *local* if it is contained in a ball, and a *core* if it is isotopic to $\{x\} \times S^1$. An embedding $\varphi\colon D^2 \times S^1 \hookrightarrow S^3$ is *trivial* (or *unknotted*) if the image of a core is a trivial knot.

Proposition 11.2.9. *A knot $K \subset S^3$ is satellite $\Longleftrightarrow$ it is the image of a knot $K' \subset D^2 \times S^1$ which is neither local nor a core, along a non-trivial embedding $\varphi\colon D^2 \times S^1 \hookrightarrow S^3$.*

Proof. A knot $K = \varphi(K')$ constructed in this way is satellite, because the torus $\varphi(S^1 \times S^1)$ is essential in the complement of K: it is incompressible (on one side because φ is non-trivial, and on the other because K' is not local) and not ∂-parallel (because K' is not a core).

Conversely, if K is a satellite knot then its complement contains an essential torus $T \subset S^3$. As every torus in S^3, the torus T bounds a solid torus. Since T is essential, the knot K is contained in this solid torus in a non-local and non-core way. Moreover the solid torus is knotted, otherwise T would be compressible on the other side. $\square$

The non-trivial embedding φ sends the core curve of $D^2 \times S^1$ to some non-trivial knot $H \subset S^3$ called the *companion* of K: we should think of K as orbiting as a "satellite" around its companion H, with orbit path K'. When K' is contained in the boundary of the solid torus $D^2 \times S^1$ the satellite knot K is called a *cable* knot. Some examples are in Figure 11.9.

11.2.7. Simple complements. Recall from Section 9.4.7 that a compact three-manifold is simple if it contains no essential sphere, disc, torus, and annulus. Which knots have a simple complement? To answer to this question, we need a general lemma on three-manifolds that identifies the few cases where there are essential annuli without essential tori.

Lemma 11.2.10. *Let M be irreducible and ∂-irreducible, with boundary consisting of tori. The manifold M contains no essential tori but contains some essential annuli ⟺ it is diffeomorphic to one of the following:*

$$\big(D, (p_1, q_1), (p_2, q_2)\big), \quad \big(A, (p, q)\big), \quad P \times S^1$$

with $p_1, p_2 \geqslant 2$. Here P is a pair of pants.

Proof. Propositions 10.4.9 and 10.4.11 imply that the Seifert manifolds listed contain vertical essential annuli, but not essential tori.

Conversely, let M contain an essential annulus A. Suppose that A connects two distinct boundary tori T, T' of M. A regular neighbourhood of $T \cup T' \cup A$ is diffeomorphic to $P \times S^1$, and its boundary contains a third torus $T'' \subset M$. Since T'' cannot be essential, it is either boundary parallel or bounds a solid torus in M, and M is diffeomorphic respectively to $P \times S^1$ or a Dehn filling of it. In the latter case, the Dehn filling is not fibre-parallel because M is ∂-reducible (use Corollary 10.3.43), hence we get a Seifert manifold of type $\big(A, (p, q)\big)$.

If A connects one boundary component to itself we conclude similarly and may also get $M = \big(D, (p_1, q_1), (p_2, q_2)\big)$.   □

We can now state and prove an important trichotomy on knots.

Proposition 11.2.11. *Every knot $K \subset S^3$ is either a torus knot, a satellite knot, or has a simple complement.*

Proof. We only need to prove that if the complement M contains an essential annulus and no essential tori then K is a torus knot. The previous lemma gives $M = \big(D, (p_1, q_1), (p_2, q_2)\big)$, and to get S^3 back we must have a Dehn filling $S^3 = \big(S^2, (p_1, q_1), (p_2, q_2), (1, n)\big)$. In particular K is isotopic to a fibre and hence contained in a vertical Heegaard torus for S^3. Heegaard tori are standard by Corollary 11.2.2, and hence K is a torus knot.   □

The relevance of this proposition will be magnified later, after introducing geometrisation. Note that in particular all the composite knots are satellite knots: where is the essential torus?

11.3. Dehn surgery

While a Dehn *filling* consists of attaching a solid torus to a boundary component, a Dehn *surgery* is a two-step operation that consists of drilling a tubular neighbourhood of a knot and then re-gluing it via a different map.

Using this fundamental operation we can describe *every* closed three-manifold via some reasonable and combinatorial drawing on the plane: a link diagram with some rational numbers attached to its components. Such a combinatorial description has a strong four-dimensional flavour and is called a *Kirby diagram*.

In this section we introduce some basic knot theory concepts: longitudes, Seifert surfaces, Dehn surgery, and the Lickorish-Wallace Theorem which asserts that every closed orientable 3-manifold is obtained by surgerying some link in S^3.

11.3.1. Canonical longitudes. The tubular neighbourhood N of a knot $K \subset S^3$ is a solid torus. As usual with solid tori, a *meridian* is a simple closed curve $m \subset \partial N$ bounding a disc in N and a *longitude* is any other simple closed curve l such that m and l generate $H_1(\partial N, \mathbb{Z})$. The meridian m is unique up to sign, but the longitude l is not: if l is a longitude, then $l + km$ also is for any $k \in \mathbb{Z}$.

The purpose of this section is to define a canonical longitude.

Proposition 11.3.1. *Let $L \subset S^3$ be a link with k components and M its complement. We have $H_1(M, \mathbb{Z}) = \mathbb{Z}^k$, generated by the k meridians.*

Proof. Let $N = N_1 \sqcup \ldots \sqcup N_k$ be the solid tori neighbourhoods of L and $T_i = \partial N_i$. The Mayer–Vietoris sequence on $S^3 = M \cup N$ gives

$$0 \longrightarrow H_1(T_1 \sqcup \ldots \sqcup T_k) \longrightarrow H_1(M) \oplus H_1(N_1 \sqcup \ldots \sqcup N_k) \longrightarrow 0$$

since $H_2(S^3) = H_1(S^3) = 0$. The equalities $H_1(T_i) = \mathbb{Z} \times \mathbb{Z}$ and $H_1(N_i) = \mathbb{Z}$ imply that $H_1(M) = \mathbb{Z}^k$. The group $H_1(T_i)$ is generated by (m_i, l_i) and m_i goes to zero in $H_1(N_i)$. Therefore the meridians $m_1, \ldots, m_k$ go to generators of $H_1(M)$. $\square$

Corollary 11.3.2. *Let $K \subset S^3$ be a knot and M be its complement. A unique (up to sign) longitude $l \subset \partial M$ vanishes in $H_1(M, \mathbb{Z})$.*

Proof. In the map $\mathbb{Z} \times \mathbb{Z} = H_1(\partial M) \to H_1(M) = \mathbb{Z}$ the meridian m goes to a generator, hence the kernel is generated by a longitude l. $\square$

We call l the *canonical longitude* of K. The torus $T = \partial M$ is hence equipped with a canonical basis (m, l) for $H_1(T, \mathbb{Z})$; we orient m and l as shown in Figure 11.6, like a right-hand screw. The pair (m, l) is well-defined up to reversing both their signs.

11.3.2. Seifert surfaces. We now show that the canonical longitude has a concrete geometric interpretation.

Definition 11.3.3. A *Seifert surface* for a knot $K \subset S^3$ is any orientable connected compact surface $S \subset S^3$ with $\partial S = K$.

Figure 11.10. A Seifert surface for the trefoil knot: it is a punctured torus.

See an example in Figure 11.10. Every Seifert surface S determines a longitude l for K: pick a small tubular neighbourhood N of K and set $l = S \cap \partial N$. The same knot K has plenty of non-isotopic Seifert surfaces, but luckily these all induce the same longitude:

Proposition 11.3.4. *Every knot K has a Seifert surface S. Every Seifert surface for K induces the canonical longitude l.*

Proof. Let M be the complement of K. Let S be a surface representing a generator of $H_2(M, \partial M) = H^1(M) = \mathbb{Z}$. The long exact sequence

$$ \ldots \longrightarrow H_2(M, \partial M) \longrightarrow H_1(\partial M) \longrightarrow H_1(M) \longrightarrow \ldots $$

implies that $[S]$ is mapped to a non-trivial primitive element $\alpha \in H_1(\partial M) = \mathbb{Z} \times \mathbb{Z}$ that is trivial in $H_1(M)$. Therefore $[\partial S] = \alpha = [l]$ and ∂S consists of an odd number of parallel copies of l and some homotopically trivial simple closed curves in $T = \partial M$. The homotopically trivial components may be eliminated by isotoping them inside M and capping them with discs; if the parallel copies are more than 1, since their signed sum is 1, there must be two of them that are close and with opposite signs, that can be canceled by isotoping them inside M and capping them with an annulus. At the end we get $\partial S = K$. $\qquad\square$

The *Seifert genus* of K is the minimum genus of a Seifert surface.

Proposition 11.3.5. *The unknot is the only knot with Seifert genus 0.*

Proof. A knot has genus zero $\iff$ it bounds a disc. $\qquad\square$

Figure 11.10 shows that the trefoil knot has genus one.

11.3.3. Dehn surgery. Let $L \subset M$ be a link with k components in an orientable 3-manifold M. A *Dehn surgery* on L is a Dehn filling of the complement of L. That is, it is a two-step operation that consists of:

(1) (*drilling*) the removal of small open tubular neighbourhoods of L, that creates new boundary tori $T_1, \ldots, T_k$;
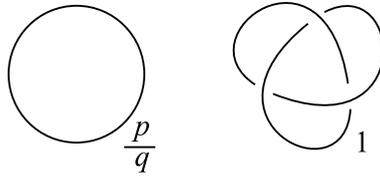(2) (*filling*) a Dehn filling of the new boundary tori $T_1, \ldots, T_k$.

Figure 11.11. Two Kirby diagrams describing the lens space
$L(p, q)$ and the Poincaré homology sphere $\Sigma(2, 3, 5)$.

We remove the tubular neighbourhoods of L and glue them back differently:
this explains the use of the word *surgery*. The outcome is a new manifold
N with the same boundary of M, that is often not diffeomorphic to M.

The surgered manifold N is determined by the slopes in $T_1, \ldots, T_k$
that are killed by the Dehn filling, see Section 10.1.1. When $M = S^3$,
every torus T_i is equipped with a canonical basis m_i, l_i, the slope is of
the form $\pm(p_i m_i + q_i l_i)$ and is hence determined by the rational number
$\frac{p_i}{q_i} \in \mathbb{Q} \cup \{\infty\}$. The full surgery on the link is comfortably encoded by
assigning the number $\frac{p_i}{q_i}$ to the i-th component of L, for every $i = 1, \ldots, k$.

The result of a Dehn surgery along $L \subset S^3$ is a closed orientable
3-manifold. The slope $\infty = \frac{1}{0}$ indicates the meridian m_i.

Proposition 11.3.6. *An ∞-surgery on a knot K has no effect.*

Proof. It consists of removing a solid torus neighbourhood of K and
regluing it back with the same map.   □

A *Kirby diagram* is a link diagram on the plane with a rational number
$\frac{p_i}{q_i}$ assigned to each component. Such a diagram defines a Dehn surgery
and hence a closed orientable three-manifold. Some examples are shown
in Figure 11.11.

Exercise 11.3.7. The $\frac{p}{q}$-surgery on the unknot yields $L(p, q)$.

The slope 0 indicates the canonical longitude l_i. More generally, the
integer $n_i \in \mathbb{Z}$ encodes the longitude $n_i m_i + l_i$. The surgery is *integral* if
the coefficients $\frac{p_i}{q_i}$ are all integers.

Remark 11.3.8. The notion of *integral* Dehn surgery exists for any
link $L \subset M$ in any 3-manifold M: a Dehn surgery is integral if the killed
slopes are longitudes of the previously removed solid tori. However on a
generic M there is no rule for choosing a canonical longitude and hence to
transform slopes into numbers.

Let N be obtained by Dehn surgery on $K \subset S^3$ with coefficient $\frac{p}{q}$.

Proposition 11.3.9. *The surgered manifold N has $H_1(N, \mathbb{Z}) = \mathbb{Z}/_{p\mathbb{Z}}$.*

Proof. Let M be the complement of the knot K. We know that the meridian m generates $H_1(M) = \mathbb{Z}$ while the longitude l is zero there. The Dehn filling kills the element $pm + ql = pm$. □

Recall that a *homology sphere* is a closed 3-manifold M having the same integral homology as S^3, that is with trivial $H_1(M, \mathbb{Z})$.

Corollary 11.3.10. *If the coefficient is $\frac{1}{q}$ the surgered manifold N is a homology sphere.*

We have found a simple method to construct many homology spheres.

11.3.4. Torus knots. As a first example, we study the effect of a Dehn surgery on torus knots. We suppose $p, q \geqslant 1$.

Proposition 11.3.11. *The $\frac{t}{u}$-Dehn surgery on a (p, q)-torus knot gives*

$$(11) \qquad \left(S^2, (p, r), (q, s), (t - pqu, u)\right) \quad \text{if } \frac{t}{u} \neq pq,$$

$$(12) \qquad \qquad L(p, r)\#L(q, s) \quad \text{if } \frac{t}{u} = pq$$

where (r, s) is any pair with $ps + qr = 1$.

Proof. Recall from the proof of Proposition 11.2.5 that the complement of K in $T \times [-1, 1]$ is diffeomorphic to $P \times S^1$. The meridian m of K is isotopic to a component of $\partial P \times \{pt\}$, so the canonical longitude l of K is isotopic to $l' + km$ for some $k \in \mathbb{Z}$, where l' is isotopic to $\{pt\} \times S^1$, *i.e.* it is the framing induced by T.

By Proposition 11.2.5 the Dehn filling is $\left(S^2, (p, r), (q, s), (t - ku, u)\right)$ and we need to determine k. Recall that if $\frac{t}{u} = \frac{0}{1}$ the manifold has infinite cyclic homology, hence it is a Seifert manifold with $e = \frac{1}{pq} - \frac{1}{k} = 0$ by Proposition 10.3.15: we deduce that $k = pq$.

When $t - pqu = 0$ the filling is fiber-parallel and Corollary 10.3.43 shows that we actually get $L(p, r)\#L(q, s)$. □

Recall from Corollary 10.3.16 that for every triple (p, q, n) of pairwise coprime natural numbers $p, q, n \geqslant 2$ there is a unique Seifert homology sphere $\Sigma(p, q, n)$ fibering over the orbifold (S^2, p, q, n).

Corollary 11.3.12. *Fix two coprime $p, q \geqslant 2$. The Seifert homology sphere obtained by $\frac{1}{u}$-surgery on the (p, q)-torus knot is $\Sigma(p, q, |pqu - 1|)$.*

In particular the 1-surgery on the trefoil knot (seen as a $(2, 3)$-torus knot) depicted in Figure 11.11 produces the Poincaré homology sphere $\Sigma(2, 3, 5)$.

Exercise 11.3.13. *Every homology sphere of type $\Sigma(2, 3, k)$ may be obtained via a Dehn surgery of the trefoil knot.*

Note that when $t - pqu = \pm 1$ the Seifert manifold (11) has only two singular fibres and is hence a lens space.
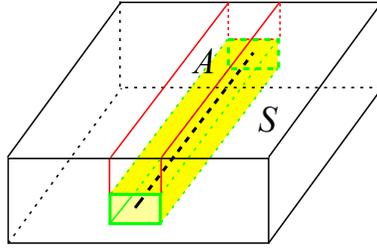
Figure 11.12. We push γ in a collar for ∂H₂ and drill a solid torus neighbourhood, drawn here as a yellow parallelepiped.

Corollary 11.3.14. *Fix two coprime $p, q \geqslant 2$ and $n \geqslant 1$. When $\frac{t}{u} = pq \pm \frac{1}{n}$ the $\frac{t}{u}$-Dehn surgery on a (p, q)-torus knot produces the lens space*

$$L(t, uq^2) = L(npq \pm 1, nq^2).$$

Proof. The Seifert manifold (11) is

$$
\begin{aligned}
\left(S^2, (p, r), (q, s), (1, \pm n)\right) &= \left(S^2, (p, r), (q, s \pm nq)\right) \\
&= L(p(s \pm nq) + rq, (-q)(s \pm nq) + sq) \\
&= L(1 \pm npq, \mp nq^2)
\end{aligned}
$$

using Exercise 10.3.6.   □

We note that we never get $S^2 \times S^1$ or S^3 via non-trivial surgeries on non-trivial torus knots. The lens space with smallest fundamental group that we may get is $L(5, 1)$, which arises from a 5-Dehn surgery on the trefoil knot.

11.3.5. The Lickorish-Wallace theorem. We now prove that the Dehn surgery construction is as general as possible.

Theorem 11.3.15 (Lickorish-Wallace theorem). *Every closed orientable 3-manifold can be described via an integral Dehn surgery along a link $L \subset S^3$.*

Proof. Let M be a closed orientable 3-manifold. Pick a Heegaard splitting $M = H_1 \cup_\psi H_2$ where H_1 and H_2 are genus-g handlebodies and $\psi \colon \partial H_1 \to \partial H_2$ is a diffeomorphism. We fix an identification of both H_1 and H_2 with a model handlebody H, so that ψ can be interpreted as an element of the mapping class group $\mathrm{MCG}(S)$ of the genus-g surface $S = \partial H$.

Example 11.1.5 shows that the three-sphere also decomposes as $S^3 = H_1 \cup_\varphi H_2$ for some $\varphi \in \mathrm{MCG}(S)$. Theorem 6.5.9 says that $\psi \circ \varphi^{-1}$ is a composition of Dehn twists

$$\psi \circ \varphi^{-1} = T_{\gamma_k}^{\pm 1} \circ \ldots \circ T_{\gamma_1}^{\pm 1}$$

along some curves $\gamma_i \subset S$. Set $M_i = H_1 \cup_{\psi_i} H_2$ with

$$\psi_i = T_{\gamma_i}^{\pm 1} \circ \ldots \circ T_{\gamma_1}^{\pm 1} \circ \varphi.$$

We have $M_0 = S^3$ and $M_k = M$. We prove that M_i can be described via an integral Dehn surgery along a i-components link in S^3 by induction on i. To obtain that it suffices to check that M_{i+1} can be obtained from M_i via integral Dehn surgery along a knot. We have

$$M_i = H_1 \cup_{\psi_i} H_2, \qquad M_{i+1} = H_1 \cup_{T_{\gamma_{i+1}}^{\pm 1} \circ \psi_i} H_2.$$

Now fix a collar of $S = \partial H_2$ in H_2 and push γ_{i+1} inside the collar as in Figure 11.12. Drill from H_2 a solid torus around γ_{i+1} (a yellow parallelepiped in the figure) to get a submanifold $H_2^{\text{drill}} \subset H_2$. We see $T_{\gamma_{i+1}}^{\pm 1}$ as a Dehn twist supported in the annulus $A \subset S$ drawn in the figure.

The Dehn twist supported on A extends product-wise to the solid torus $A \times [0, 1]$ lying between A and the drilled yellow parallelepiped, and extends trivially to a self-diffeomorphism $T \colon H_2^{\text{drill}} \to H_2^{\text{drill}}$ such that $T|_S = T_{\gamma_{i+1}}^{\pm 1}$. We define

$$M_j^{\text{drill}} = H_1 \cup_{\psi_j} H_2^{\text{drill}}$$

for $j = i, i+1$. Since $T|_S = T_{\gamma_{i+1}}^{\pm 1}$ the map T extends to a diffeomorphism

$$T \colon M_i^{\text{drill}} \longrightarrow M_{i+1}^{\text{drill}}.$$

Therefore M_{i+1} is obtained from M_i by Dehn surgery along γ_{i+1}. The surgery is integral since T sends a meridian of the drilled yellow solid torus to a longitude. The proof is complete. □

The theorem shows that every closed orientable three-manifold can be constructed using Kirby diagrams. As for knots and many other combinatorial descriptions of topological objects, it is important to note that many different Kirby diagrams may describe the same three-manifold, and in general it is hard to tell, given two diagrams, whether they define the same manifold or not.

11.3.6. Four-manifolds. The theorem of Lickorish and Wallace has a nice four-dimensional interpretation, which shows in particular that every closed orientable three-manifold is the boundary of a simply-connected four-manifold.

Recall from Section 1.7.5 that in dimension four a two-handle $H = D^2 \times D^2$ is attached to an orientable four-manifold W via an embedding $\psi \colon S^1 \times D^2 \to \partial W$. After attaching the handle H to W, the boundary of W changes by substituting the solid torus $\psi(S^1 \times D^2)$ with another solid torus $D^2 \times S^1$: in other words, it changes by Dehn surgery. The Dehn surgery is integral because the meridians of the two solid tori intersect in one points.

Conversely, every integral Dehn surgery on a knot $K \subset \partial W$ may be interpreted as the effect of attaching some 2-handle: it suffices to interpret

the drilled and filled solid tori as horizontal and vertical boundaries of a four-dimensional two-handle.

With such an interpretation, we may consider S^3 as the boundary of D^4 and see the integral Dehn surgery on a link $L \subset S^3$ as the result of attaching some two-handles to D^4. The Lickorish-Wallace theorem implies the following.

Corollary 11.3.16. *Every closed orientable three-manifold is the boundary of a simply-connected compact four-manifold.*

Proof. Every closed orientable three-manifold is the result of an integral surgery on a link $L \subset S^3$ and is hence the boundary of a four-manifold obtained by attaching some two-handles to D^4. Every such manifold is simply-connected by Van Kampen's theorem. □

This theorem reveals in particular that it is fairly easy to construct plenty of simply connected four-manifolds with boundary. Note that the simply connected four-manifolds constructed in this way are not contractible: indeed the following proposition shows that they have a non-trivial second homology group.

Proposition 11.3.17. *A four-manifold with one 0-handle and k 2-handles is homotopy equivalent to a bouquet of k two-spheres.*

Proof. We first deformation retract the two-handles over their core discs, and then we shrink the whole 0-handle to a point. □

To construct contractible four-manifolds, we need to employ 1-handles.

11.3.7. Three-manifolds with boundary. The Lickorish-Wallace Theorem extends to compact manifolds with boundary, in the appropriate way.

Proposition 11.3.18. *Every compact orientable 3-manifold M with boundary is obtained from S^3 as follows: pick $L \sqcup C \subset S^3$ where L is a link and C is a 1-complex; remove an open regular neighbourhood of C and perform an integral surgery on L.*

Proof. Cap off the boundary of M by adding handlebodies H_g, to get a closed manifold M'. Choose a 1-complex $C \subset M'$ whose regular neighbourhood consists of these handlebodies, so that by removing C we get M back.

Now M' is obtained from S^3 via integral surgery along some link L, and by general position we may suppose that C is disjoint from the cores of the surgered solid tori, so we can see both C and L disjointly in S^3. □

If ∂M consists of tori we may suppose that C is also a link.

11.4. Surface bundles

Seifert manifolds have base orbifolds of dimension 2 and smooth fibres of dimension 1. We now introduce a complementary construction where the base orbifolds have dimension 1 and the smooth fibres have dimension 2. These are called *surface bundles*.

11.4.1. Surface bundles. A *surface bundle over S^1* is a fibre bundle $M \to S^1$ of a compact orientable 3-manifold M (possibly with boundary) over S^1, whose fibre Σ is a connected compact orientable surface. If M has boundary, then Σ also has, and ∂M consists of tori fibering over S^1.

Proposition 11.4.1. *Every surface bundle over S^1 is constructed by taking $\Sigma \times [0, 1]$ and glueing $\Sigma \times 0$ to $\Sigma \times 1$ via an orientation-preserving diffeomorphism ψ.*

Proof. One such glueing clearly gives rise to a surface bundle over S^1. Conversely, by cutting a surface bundle over S^1 along a fibre we get a surface bundle over the interval, which is a product $\Sigma \times [0, 1]$. □

The diffeomorphism ψ is the *monodromy* of the surface bundle M_ψ. Since isotopic glueings produce diffeomorphic manifolds, the three-manifold M_ψ depends only of the class of ψ in the mapping class group MCG(Σ) on Σ. More than that, it actually depends only on its conjugacy class:

Proposition 11.4.2. *If ψ and ϕ are conjugate in MCG(Σ), then M_ψ and M_ϕ are diffeomorphic.*

Proof. The diffeomorphism $g \colon \Sigma \to \Sigma$ that conjugates them extends to $\Sigma \times [0, 1]$ and gives a diffeomorphism $M_\psi \to M_\phi$. □

The manifolds M_ψ and $M_{\psi^{-1}}$ are orientation-reversingly diffeomorphic.

11.4.2. Properties. We now start to investigate the topological properties of surface bundles. Let $M \to S^1$ be a surface bundle with fibre Σ.

Exercise 11.4.3. The maps $\Sigma \to M \to S^1$ induce an exact sequence
$$0 \longrightarrow \pi_1(\Sigma) \longrightarrow \pi_1(M) \longrightarrow \pi_1(S^1) \longrightarrow 0.$$

This implies in particular that $\pi_1(M)$ surjects onto $\mathbb{Z}$ and therefore:

Corollary 11.4.4. *We have $b_1(M) \geqslant 1$.*

In other terms, the surface fibre Σ is non-separating and hence $[\Sigma] \in H_2(M, \partial M)$ is non-trivial (and has infinite order).

Note that there is an obvious degree-n regular covering $M_{\psi^n} \to M_\psi$ for every n and an infinite regular covering $\Sigma \times \mathbb{R} \to M_\psi$ induced by the normal subgroup $\pi_1(\Sigma) \lhd \pi_1(M)$.

Proposition 11.4.5. *The fibre Σ is an essential surface. If $\chi(\Sigma) > 0$ then M is diffeomorphic to $D \times S^1$ or $S^2 \times S^1$. If $\chi(\Sigma) \leqslant 0$ the universal cover of $\text{int}(M)$ is $\mathbb{R}^3$ and M is Haken.*

Proof. If $\chi(\Sigma) > 0$ then MCG(Σ) is trivial and we are done, so we suppose $\chi(\Sigma) \leqslant 0$. The fibre Σ is incompressible because $\pi_1(\Sigma)$ injects, and is also ∂-incompressible by a doubling argument (the double DM fibres to S^1 with incompressible fibre $D\Sigma$, hence Σ is ∂-incompressible). The fibre is clearly not ∂-parallel, hence it is essential.

The manifold M is covered by $\Sigma \times \mathbb{R}$, whose interior is covered by $\mathbb{R}^2 \times \mathbb{R} = \mathbb{R}^3$: hence M is irreducible. It is also ∂-irreducible because its double also fibres and hence is irreducible. Therefore M is Haken.   □

11.4.3. Semi-bundles. The Seifert fibrations are circle bundles over two-dimensional orbifolds, and likewise it is natural to consider surface bundles over 1-orbifolds. The compact 1-orbifolds are S^1 and the closed segment with mirrored endpoints. We have already considered the S^1 case and we now define some surface bundles over the closed segment, called *semi-bundles*.

Let Σ be a non-orientable surface, with orientable double cover $p\colon \tilde{\Sigma} \to \Sigma$ and deck transformation τ that gives $\Sigma = \tilde{\Sigma}/_\tau$. We have

$$\Sigma \mathbin{\tilde{\times}} (-1, 1) = \big(\tilde{\Sigma} \times (-1, 1)\big)/_{(\tau, \iota)}$$

with $\iota(x) = -x$. A *local semi-bundle* is the map $\Sigma \mathbin{\tilde{\times}} (-1, 1) \to [0, 1)$ that sends (p, x) to $|x|$. The fibre over 0 is Σ, that over $x \in (0, 1)$ is $\tilde{\Sigma}$.

A *semi-bundle* $M \to [-1, 1]$ is a map which is a local semi-bundle when restricted to $[-1, 1 - \varepsilon]$ and $(\varepsilon, 1]$. The fibre over ± 1 is Σ and the fibre of $x \in (-1, 1)$ is $\tilde{\Sigma}$. We should think about this object as a surface bundle over the segment orbifold $[-1, 1]$ with mirror points ± 1.

Exercise 11.4.6. Every semi-bundle is constructed by gluing two copies of $\Sigma \mathbin{\tilde{\times}} [-1, 1]$ along their boundaries via some diffeomorphism.

Set $I = [-1, 1]$. Let $M \to I$ be a semi-bundle with fibres Σ and $\tilde{\Sigma}$. Note that $I = \mathbb{R}/_\Gamma$ where $\Gamma < \text{Isom}(\mathbb{R})$ is generated by the reflections at the points ± 1. The orbifold fundamental group of I is

$$\pi_1(I) = \Gamma = \mathbb{Z}/_{2\mathbb{Z}} * \mathbb{Z}/_{2\mathbb{Z}}.$$

Exercise 11.4.7. The maps $\tilde{\Sigma} \to M \to I$ induce an exact sequence

$$0 \longrightarrow \pi_1(\tilde{\Sigma}) \longrightarrow \pi_1(M) \longrightarrow \pi_1(I) \longrightarrow 0.$$

Proposition 11.4.8. *If $\chi(\Sigma) > 0$ then M is diffeomorphic to $\mathbb{RP}^2 \mathbin{\tilde{\times}} S^1$. If $\chi(\Sigma) \leqslant 0$ the orientable fibre $\tilde{\Sigma}$ is essential, the universal cover of $\text{int}(M)$ is $\mathbb{R}^3$, and M is Haken.*

Proof. Same proof as in the standard bundle case.   □

Exercise 11.4.9. An orientable 3-manifold M has a (semi-)bundle structure if and only if there is an orientable $\Sigma \subset M$ that cuts M into interval bundles.

Let $M \to I$ be a semi-bundle. We may pull-back the semi-bundle along the orbifold double cover $S^1 \to I$ and get an ordinary surface bundle $\tilde{M} \to S^1$, so that the following diagram commutes:

$$
\begin{array}{ccc}
\tilde{M} & \dashrightarrow & M \\
\downarrow & & \downarrow \\
S^1 & \longrightarrow & I
\end{array}
$$

Every semi-bundle is thus covered by an ordinary bundle.

11.4.4. Seifert manifolds. We now classify the Seifert manifolds that have a surface (semi-)bundle structure.

Proposition 11.4.10. *A Seifert manifold M has a (semi-)bundle structure if and only if one of the following holds:*

- $\partial M \neq \varnothing$,
- $e(M) = 0$,
- $M = \big(T, (1, e)\big)$ *or* $\big(K, (1, e)\big)$,
- $M = \big(S^2, (2, 1), (2, 1), (2, 1), (2, 2q + 1)\big)$,
- $M = \big(\mathbb{RP}^2, (2, 1), (2, 2q + 1)\big)$.

The manifolds in the last two lines occur only as semi-bundles.

Proof. If $\partial M \neq \varnothing$ or $e(M) = 0$ then $M \to S$ has an orientable section Σ, which cuts M into interval bundles: hence Σ is a fibre of a (semi-)bundle (see Exercise 11.4.9).

Let S be one of the orbifolds

$$
T, \quad K, \quad (S^2, 2, 2, 2, 2), \quad (\mathbb{RP}^2, 2, 2).
$$

It contains a circle that splits S into one or two annuli, Möbius strips, or $(D, 2, 2)$. Every circle fibering over these pieces is an interval bundle by Proposition 10.3.33. Therefore every fibering $M \to S$ contains a vertical torus that splits M into interval bundles and is hence a (semi-)bundle.

Conversely, if M has a (semi-)bundle structure the orientable fibre is essential and is hence either horizontal or vertical: we get one of the types listed. $\qquad\square$

11.4.5. Torus bundles. A *torus bundle* is of course a surface bundle $M \to S^1$ with fibre a torus T. We fix a basis for $\pi_1(T)$, so that $\mathrm{MCG}(T) = \mathrm{SL}_2(\mathbb{Z})$. By what said above, every matrix $A \in \mathrm{SL}_2(\mathbb{Z})$ defines a torus bundle M_A with monodromy A. We want to understand when M_A is a Seifert manifold.

Exercise 11.4.11. For every $e \in \mathbb{Z}$ there are diffeomorphisms

$$
M_{\left(\begin{smallmatrix} 1 & e \\ 0 & 1 \end{smallmatrix}\right)} \cong \big(T, (1, e)\big), \quad M_{\left(\begin{smallmatrix} -1 & e \\ 0 & -1 \end{smallmatrix}\right)} \cong \big(K, (1, e)\big).
$$

The classification of torus bundles reduces to linear algebra.

Proposition 11.4.12. *Two torus bundles M_A and $M_{A'}$ are diffeomorphic if and only if A' is conjugate to $A^{\pm 1}$ in $\mathrm{SL}_2(\mathbb{Z})$.*

Proof. Let $M_A \cong M_{A'}$ be a diffeomorphism. Consider the tori T and T' of the two fibrations both inside M_A and minimise their transverse intersection.

If $T \cap T' = \varnothing$, by Proposition 9.3.18 the tori are parallel and the two fibrations are isotopic, so A' is conjugate to $A^{\pm 1}$. Otherwise by the same proposition T' decomposes into fibered annuli in the product $M_A \setminus T$. One such fibered annulus identifies a circle $\gamma \subset T$ preserved by A. Thus A is conjugate to $\left(\begin{smallmatrix} 1 & e \\ 0 & 1 \end{smallmatrix}\right)$ or $\left(\begin{smallmatrix} -1 & e \\ 0 & -1 \end{smallmatrix}\right)$, and A' is conjugate to $\left(\begin{smallmatrix} 1 & e' \\ 0 & 1 \end{smallmatrix}\right)$ or $\left(\begin{smallmatrix} -1 & e' \\ 0 & -1 \end{smallmatrix}\right)$ for the same reason. Exercise 11.4.11 implies that $A' = A^{\pm 1}$. □

Exercise 11.4.13. A matrix $A \in \mathrm{SL}_2(\mathbb{Z})$ has finite order if and only if $A = \pm I$ or $|\mathrm{tr}A| < 2$. Every finite-order A is conjugate to one of

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \begin{pmatrix} -1 & 1 \\ -1 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad \begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix}$$

or their inverses. These matrices have order 1, 2, 3, 4, 6.

Hint. Use Proposition 2.3.9. □

We can easily determine whether a torus bundle is Seifert by looking at its monodromy A.

Proposition 11.4.14. *Let $M = M_A$ be a torus bundle with monodromy $A \neq \pm I$. The following holds:*

- *if $|\mathrm{tr}A| < 2$ then M is a Seifert manifold with $e = 0$ and $\chi = 0$,*
- *if $|\mathrm{tr}A| = 2$ then M is a Seifert manifold with $e \neq 0$ and $\chi = 0$,*
- *if $|\mathrm{tr}A| > 2$ then M is not a Seifert manifold.*

Proof. Consider $T \times [-1, 1]$ foliated by lines $\{x\} \times [-1, 1]$. The foliation extends to M_A. If A has finite order, then M_A is finitely covered by $M_I = T \times S^1$ and hence all fibers are compact. Therefore M_A is Seifert fibered and covered by $T \times S^1$, and we get $e = \chi = 0$.

If $|\mathrm{tr}A| = 2$ then A is conjugate to $\left(\begin{smallmatrix} \pm 1 & e \\ 0 & \pm 1 \end{smallmatrix}\right)$ and we use Exercise 11.4.11. Proposition 11.4.10 easily shows that all the Seifert manifolds that are torus bundles are realised with $|\mathrm{tr}A| \leqslant 2$, hence if $|\mathrm{tr}A| > 2$ the manifold M is not Seifert by Proposition 11.4.12. □

When $|\mathrm{tr}A| > 2$ we say that the monodromy A is *Anosov*.

11.4.6. Bundles with $\chi(\Sigma) < 0$. Proposition 11.4.12 does not extend to surface bundles with $\chi(\Sigma) < 0$; indeed it may happen that non-conjugate monodromies in $\mathrm{MCG}(\Sigma)$ give rise to diffeomorphic manifolds and understanding when this happens is a hard problem.

Proposition 11.4.14 extends nevertheless and reflects the trichotomy of mapping classes. Let Σ be a closed orientable surface with $\chi(\Sigma) < 0$.

Recall from Section 8.4 that every element $\psi \in \mathrm{MCG}(\Sigma)$ is either finite order, reducible, or pseudo-Anosov.

Proposition 11.4.15. *Let M_ψ be a surface bundle with fibre Σ and monodromy $\psi \in \mathrm{MCG}(\Sigma)$. The following holds:*

- *if ψ has finite order, then M_ψ is Seifert with $\chi < 0$ and $e = 0$,*
- *if ψ is reducible, then M_ψ contains an essential torus,*
- *if ψ is pseudo-Anosov, then M_ψ is simple and not Seifert.*

Proof. Same proof as Proposition 11.4.14. If ψ has finite order, it is an isometry for some hyperbolic metric on Σ, and the line fibration of $\Sigma \times [-1, 1]$ glues to a Seifert fibration for M_ψ.

If ψ is reducible there are some disjoint essential simple closed curves $\gamma_1, \ldots, \gamma_k$ with $\psi(\gamma_i) = \gamma_{i+1}$ cyclically; by gluing the annuli $\gamma_i \times [-1, 1]$ we get a torus $T \subset M_\psi$. It is essential because by cutting along it we still get a fibration over S^1 with fibers having $\chi \leqslant 0$.

If ψ is pseudo-Anosov there are no essential tori $T \subset M_\psi$, for by minimising $T \cap \Sigma$ then $M_\psi \setminus T$ would consist of essential annuli of type $\gamma \times [-1, 1]$ and hence ψ would be reducible. The manifold M_ψ is not Seifert because the fibre Σ would become a horizontal surface: then M would be covered by $\Sigma \times S^1$ and hence ψ would be of finite order. $\square$

11.5. JSJ decomposition

In Chapter 9 we cut every closed three-manifold along spheres, and it is now time to decompose it further along tori. This two-steps cutting operation is called the *JSJ decomposition* of the manifold, after the names of Jaco, Shalen, and Johansson who discovered it in the mid 1970s.

The core of this decomposition is the existence of a canonical set of disjoint essential tori, unique up to isotopy.

11.5.1. Canonical torus decomposition. Let M be an orientable irreducible and ∂-irreducible compact 3-manifold with (possibly empty) boundary consisting of tori. How can we define a canonical set of disjoint essential tori in M? The answer is not obvious: for instance, if M is a Seifert manifold, it may contain many vertical incompressible tori and there is no canonical way to choose among them.

We will soon see that the Seifert manifolds are in fact the only possible source of ambiguity. Let

$$S = T_1 \sqcup \cdots \sqcup T_k$$

be a set of disjoint essential tori $T_i \subset \mathrm{int}(M)$. We say that S is a *torus decomposition* of M if it decomposes M into blocks that are either:

- torus (semi-)bundles,
- Seifert manifolds, or
- simple manifolds.

A torus decomposition is *minimal* if no proper subset of S is a torus decomposition. We prove here the following.

Theorem 11.5.1 (JSJ decomposition). *Let M be an orientable irreducible and ∂-irreducible compact 3-manifold with (possibly empty) boundary consisting of tori. A minimal torus decomposition for M exists and is unique up to isotopy.*

Such a minimal decomposition is called the *canonical torus decomposition* or the *JSJ decomposition* of M. The canonical torus decomposition may be empty: this holds precisely when M is itself a torus (semi-)bundle, Seifert, or simple.

Remark 11.5.2. Torus (semi-)bundles are closed: therefore if M is not itself a torus (semi-)bundle, the blocks of its canonical decomposition are either Seifert or simple.

11.5.2. Existence and uniqueness. Let M be an orientable irreducible and ∂-irreducible compact 3-manifold with (possibly empty) boundary consisting of tori. We now prove Theorem 11.5.1. Let us start by showing existence.

Proposition 11.5.3. *The manifold M has a torus decomposition.*

Proof. Let $T_1, \ldots, T_k$ be a maximal set of disjoint non-parallel essential tori in M, which exists by Corollary 9.4.8. We now prove that $S = T_1 \sqcup \cdots \sqcup T_k$ is a torus decomposition.

Suppose it is not: one block N of the decomposition is neither a (semi-)bundle, nor Seifert, nor simple. The block N is irreducible and ∂-irreducible since these properties are preserved after cutting along incompressible surfaces. Being not simple, it contains an essential annulus A or an essential torus T.

In the latter case we can add T to the family $T_1, \ldots, T_k$ and get a contradiction since S is maximal. In the former case Lemma 11.2.10 applies and N is Seifert. $\square$

Since M has a torus decomposition, it certainly has a minimal one. We now prove that it is unique.

Proposition 11.5.4. *The manifold M has a unique minimal torus decomposition up to isotopy.*

Proof. Let $S = T_1 \sqcup \cdots \sqcup T_k$ and $S' = T'_1 \sqcup \cdots \sqcup T'_{k'}$ be two minimal torus decompositions for M. We minimise their transverse intersections, so that $S \cap S'$ consists of essential circles cutting some tori into annuli.

Let T'_i be decomposed into some annuli. Each such annulus is essential in $M \setminus S$, hence it is contained in some non-simple block, *i.e.* a Seifert one. It is contained there horizontally or vertically: in the former case, the block is $\big(D, (2,1), (2,1)\big)$, $S_* \widetilde{\times} S^1$, or $A \times S^1$ with S_* the Möbius strip.

The first two blocks are diffeomorphic, and by swapping the fibration the annulus becomes vertical. The third block $T \times I$ is excluded since S is minimal.

Now all annuli in T_i' are vertical. Two consequent vertical annuli are separated by some torus T_j; since the two annuli are fibered, the fibers of the two Seifert blocks incident to T_j are isotopic: hence the two blocks glue to a bigger Seifert block and T_j can be removed, a contradiction since S is minimal.

We have shown that $S \cap S' = \varnothing$. If T_i is parallel to T_j' we superpose the two tori, cut M along $T_i = T_j'$ and proceed by induction. Now we suppose by contradiction that there is no parallelism.

Every T_i' is an essential vertical torus in a Seifert block of $M \setminus S$, and vice versa. This easily implies that all the blocks in $M \setminus S$, $M \setminus S'$, and all their intersections are Seifert! Pick one such intersection. It has a unique Seifert fibration, unless it is $K \mathbin{\tilde{\times}} I$ which may fiber in two ways. Since $\partial(K \mathbin{\tilde{\times}} I)$ is connected, one block is $K \mathbin{\tilde{\times}} I$ itself and we change the fibration on this block if necessary. Now all intersections and all blocks have unique fibrations and they all glue to a Seifert fibration for M, a contradiction. □

The proof of Theorem 11.5.1 is complete.

Remark 11.5.5. The sphere decomposition of Theorem 9.2.29 and the torus decomposition of Theorem 11.5.1 differ in two aspects: (i) the set of decomposing spheres is *not* canonical up to isotopy, while the set of tori is; (ii) on the other hand, after cutting along the spheres and capping off we get a canonical set of prime manifolds, whereas if we cut along the tori we get some canonical manifolds with toric boundaries, but there is no canonical way to cap them off.

11.5.3. Geometric decomposition. The *geometric decomposition* is a slight variation of the canonical torus decomposition that is more suited to the geometrisation perspective that we will encounter in the next chapters. It is constructed from the canonical torus decomposition S for M as follows. Whenever a block N of the torus decomposition is diffeomorphic to $K \mathbin{\tilde{\times}} I$, we substitute the torus ∂N in S with the core Klein bottle K of $K \mathbin{\tilde{\times}} I$. This substitution has the effect of deleting N from the list of blocks of the decomposition.

The geometric decomposition consists of incompressible tori and Klein bottles. One reason for preferring the geometric decomposition to the canonical torus one is that it contains no interval bundles and every Seifert block has a unique fibration up to isotopy (because we have eliminated the $K \mathbin{\tilde{\times}} I$ blocks). In particular we get the following easy criterion, whose proof is straightforward.

Proposition 11.5.6. *Let M be an orientable irreducible ∂-irreducible compact 3-manifold with (possibly empty) boundary consisting of tori. A non-empty torus decomposition of M is the geometric one if and only if:*

- *every block is simple or Seifert with $\chi < 0$,*
- *the fibrations of two adjacent Seifert blocks do not match.*

Another reason for preferring the geometric decomposition is that every Seifert block can be geometrised with finite volume, whereas $K \,\widetilde{\times}\, I$ needs infinite volume (this will be shown in the next chapter).

Geometric decompositions also behave well under finite coverings. Let $\tilde{M} \to M$ be a finite covering between two orientable irreducible and ∂-irreducible compact 3-manifolds with (possibly empty) boundary consisting of tori.

Proposition 11.5.7. *The geometric decomposition of $\tilde{M}$ is the counterimage of that of M.*

Proof. The criterion of Proposition 11.5.6 lifts from M to $\tilde{M}$. □

11.5.4. Graph manifolds. Waldhausen introduced in the 1960s a simple but non-trivial class of three-manifolds using only two blocks. Here D and P are the disc and the pair-of-pants.

Definition 11.5.8. A *graph manifold* is any orientable three-manifold that decomposes along disjoint tori into pieces diffeomorphic to $D \times S^1$ or $P \times S^1$.

One may describe any such manifold via a graph with vertices of valence 1 and 3 representing the blocks, and some appropriate 2×2-matrices labelling the edges telling the way the two incident blocks are glued. As usual, different graphs may represent the same manifold.

Exercise 11.5.9. Let M be an orientable three-manifold. The following are equivalent:

(1) M is a graph manifold,
(2) $M = M_1 \# \ldots \# M_h$ for some prime manifolds M_i whose geometric decompositions consist of Seifert manifolds or torus (semi-)bundles.

Hint. Use Corollary 10.3.44 and Proposition 11.5.6. □

11.5.5. References. Most of the arguments contained in this chapter are well-known to three-dimensional topologists and can be found in many books. A standard introduction to knots and links is Rolfsen [**50**], and much more on Dehn surgeries and their four-dimensional interpretations is contained in Gompf − Stipsicz [**23**]. A proof of the JSJ decomposition can be found in Hatcher [**26**], the original papers of Jaco − Shalen and Johannson are [**31**] and [**32**]. Fomenko − Matveev [**18**] contains a chapter devoted to graph manifolds; the original paper of Waldhausen is [**60**].

The eight geometries

We have concluded the previous chapter by defining a *geometric decomposition* of three-manifolds along sphere and tori. The reason for using this terminology is the famous *geometrisation conjecture*, proposed by Thurston in 1982 and proved by Perelman in 2003, which states that each of the blocks of the decomposition should be *geometric*, in the sense that it may be equipped with a nice Riemannian metric.

There are eight nice Riemannian metrics available in dimension three. Three of them are the constant curvature ones (hyperbolic, elliptic, and flat), while the other five are some kind of (sometimes twisted) products of low-dimensional geometries. All these metrics are *homogeneous*: distinct points have isometric neighbourhoods.

The Seifert manifolds studied in the previous chapters occupy precisely six of these eight geometries, and we analyse them in detail here.

12.1. Introduction

A connected Riemannian manifold M is *homogeneous* if for every $p, q \in M$ there is an isometry of M sending p to q, and is *isotropic* if at each point p every isometry of $T_p M$ is realised by an isometry of M. It is easy to prove that a complete isotropic manifold is also homogeneous and has constant sectional curvature: the fundamental examples of isotropic spaces are S^n, $\mathbb{R}^n$, and $\mathbb{H}^n$.

The homogeneous condition alone (without isotropy) is more relaxed and produces manifolds that may not have constant sectional curvature. We introduce here eight important homogeneous simply-connected complete Riemannian 3-manifolds:

$$S^3, \ \mathbb{R}^3, \ \mathbb{H}^3, \ S^2 \times \mathbb{R}, \ \mathbb{H}^2 \times \mathbb{R}, \ \mathrm{Nil}, \ \mathrm{Sol}, \ \widetilde{\mathrm{SL}_2}.$$

The first three manifolds are also isotropic and have constant sectional curvature, the other five are not.

Let M be one of these eight model manifolds. We say that a Riemannian 3-manifold N has a *geometric structure modelled on M* if N is locally isometric to M, that is if every point $p \in N$ has an open neighbourhood isometric to some open set in M. This implies that N is *locally homogeneous*: every two points $p, q \in N$ have isometric neighbourhoods $U(p) \cong U(q)$, both isometric to an ε-ball at any point of M.

	$\chi > 0$	$\chi = 0$	$\chi < 0$
$e = 0$	$S^2 \times \mathbb{R}$	$\mathbb{R}^3$	$\mathbb{H}^2 \times \mathbb{R}$
$e \neq 0$	S^3	Nil	$\widetilde{\mathrm{SL}_2}$

Table 12.1. The closed manifolds modelled on six geometries are precisely the six commensurable classes of Seifert manifolds, distinguished by the numbers e and χ.

If N is complete, the developing map construction of Theorem 3.1.2 applies also in this context and shows that $N = M/_\Gamma$ for some discrete group $\Gamma < \mathrm{Isom}(M)$ acting freely.

We now start a long journey through these eight geometries. The final goal of this chapter is to prove the following.

Theorem 12.1.1. *A closed orientable 3-manifold has a geometric structure modelled on one of the following six geometries:*

$$S^3, \ \mathbb{R}^3, \ S^2 \times \mathbb{R}, \ \mathbb{H}^2 \times \mathbb{R}, \ \mathrm{Nil}, \ \widetilde{\mathrm{SL}_2}$$

if and only if it is a Seifert manifold of the appropriate commensurability class, as prescribed by Table 12.1. It has a Sol *geometric structure if and only if it is a torus (semi-)bundle of Anosov type.*

It is a surprising (and maybe disappointing) fact that, despite its elegance and generality, the only known proof of this theorem available today works by investigating each geometry separately and carefully, often employing quite different techniques. We start with the elliptic case.

12.2. Elliptic three-manifolds

We start our journey by investigating elliptic 3-manifolds, that is manifolds modelled on S^3. We want to prove the following.

Theorem 12.2.1. *A closed 3-manifold M admits an elliptic metric if and only if it is a Seifert manifold with $e \neq 0$ and $\chi > 0$.*

An important ingredient of the proof is the complete classification of elliptic three-manifolds: to achieve this goal we need to study the isometries of S^3, and these are described elegantly via quaternions.

12.2.1. Unit quaternions. We write as usual a *quaternion* as

$$q = a + bi + cj + dk$$

with $a, b, c, d \in \mathbb{R}$ and $i^2 = j^2 = k^2 = ijk = -1$. Quaternions form a non-commutative algebra, identified with $\mathbb{R}^4$ by sending q to (a, b, c, d) and with $\mathbb{C}^2$ by sending q to $(a + bi, c + di)$. The norm is

$$|q| = \sqrt{a^2 + b^2 + c^2 + d^2}$$

and we have $|qq'| = |q||q'|$. The conjugate of q is

$$\bar{q} = a - bi - cj - dk$$

and we have $|q|^2 = q\bar{q}$. Unit quaternions are identified with S^3 and are closed under multiplication: this gives S^3 a Lie group structure.

Exercise 12.2.2. The centre of S^3 is $\{\pm 1\}$.

Left or right multiplication by a fixed element $q \in S^3$ is an orientation-preserving isometry of both $\mathbb{R}^4$ and S^3. We consider the homomorphism

$$\Psi \colon S^3 \times S^3 \longrightarrow \text{Isom}^+(S^3) = \text{SO}(4)$$

$$(q_1, q_2) \longmapsto \{x \longmapsto q_1 x q_2^{-1}\}$$

Proposition 12.2.3. *The homomorphism Ψ is a degree-2 covering with kernel $\{\pm(1, 1)\}$. Therefore it induces an isomorphism*

$$\text{SO}(4) = S^3 \times S^3 / {\{\pm(1,1)\}}.$$

Proof. If (q_1, q_2) lies in the kernel, by setting $x = 1$ we get $q_1 q_2^{-1} = 1$ and so $q_1 = q_2$. The general x implies that $q_1 = q_2$ lies in the centre and hence $(q_1, q_2) = \pm(1, 1)$. Since SO(4) has the same dimension 6 of $S^3 \times S^3$ and is connected, we get a covering by Proposition 1.4.21. $\square$

We now specialise Ψ to the case $q = q_1 = q_2$. The isometry

$$x \longmapsto q x q^{-1}$$

fixes 1 and hence preserves the orthogonal 3-space generated by i, j, k, which we identify with $\mathbb{R}^3$. We get a homomorphism

$$\Phi \colon S^3 \longrightarrow \text{Isom}^+(\mathbb{R}^3) = \text{SO}(3)$$

$$q \longmapsto \{x \longmapsto q x q^{-1}\}$$

Proposition 12.2.4. *The homomorphism Φ is a degree-2 covering with kernel ± 1. Therefore it induces an isomorphism*

$$\text{SO}(3) = S^3 / {\{\pm 1\}}.$$

Proof. The centre of S^3 is $\{\pm 1\}$ and SO(3) is connected and has dimension 3 like S^3, hence Φ is a covering. $\square$

The real part of $q = a + bi + cj + dk$ is of course a.

Corollary 12.2.5. *Two elements $q, q' \in S^3$ are conjugate if and only if they have the same real part.*

Proof. Conjugations are isometries that fix the real axis and hence preserve the real part; rotations in SO(3) connect any two elements with the same real part. $\square$

Corollary 12.2.6. *Every unit quaternion is conjugate to a unit complex quaternion $q = a \pm bi$, unique up to complex conjugation.*

The conjugacy classes in S^3 are the poles $+1$ and -1, and the parallel two-spheres between them. Let the *imaginary two-sphere* be the maximal two-sphere consisting of all elements with zero real part.

Corollary 12.2.7. *The only element in S^3 of order two is -1. The elements of order four form the imaginary two-sphere and are all conjugate.*

Proof. An element of order four is conjugate to a complex one $a + bi$, which must be $\pm i$ and is hence purely imaginary.   □

Exercise 12.2.8. By sending the unit quaternion $(w, z) \in \mathbb{C}^2$ to the matrix

$$\begin{pmatrix} w & z \\ -\bar{z} & \bar{w} \end{pmatrix}$$

we get a Lie group isomorphism between S^3 and $\mathrm{SU}(2)$.

Note that $\bar{q} = q^{-1}$ on unit quaternions. The inversion $q \mapsto q^{-1}$ is an orientation-reversing isometry of S^3.

12.2.2. Finite groups of quaternions. We classify the finite subgroups of the Lie group S^3. Recall from Proposition 6.2.15 that the finite subgroups of $\mathrm{SO}(3)$ up to conjugation are:

$$C_n, \quad D_{2m}, \quad T_{12} \cong A_4, \quad O_{24} \cong S_4, \quad I_{60} \cong A_5$$

with $n \geqslant 1$ and $m \geqslant 2$. Here C_n is cyclic generated by a $\frac{2\pi}{n}$-rotation, and D_{2m}, T, O, I are the orientation-preserving isometry groups of the regular m-prism, tetrahedron, octahedron (or cube), and icosahedron (or dodecahedron). The group D_{2m} is the dihedral group. The subscript always indicates the order of the group (except in the alternating A_n and symmetric S_n that have order $\frac{n!}{2}$ and $n!$).

We defined above a degree-2 covering $\Phi\colon S^3 \to \mathrm{SO}(3)$. Let

$$D_{4m}^*, \quad T_{24}^*, \quad O_{48}^*, \quad I_{120}^*$$

be the counterimages of D_{2m}, T_{12}, O_{24}, I_{60} along Φ. They are called the *binary* dihedral, tetrahedral, octahedral, and icosahedral group. We now classify the finite subgroups of S^3.

Proposition 12.2.9. *Every finite subgroup of S^3 is either binary or cyclic. The finite subgroups up to conjugation are listed in Table 12.2.*

Proof. If $G < S^3$ is finite then $\Phi(G) = C_n, D_{2m}, T, O$, or I up to conjugation. If $\Phi^{-1}(\Phi(G)) = G$ we are done (the counterimage of C_n is still cyclic: exercise). This holds precisely when $-1 \in G$.

If G has even order, it contains an order-2 element which is necessarily -1 and we are done. If G has odd order, $\Phi(G)$ has odd order and is cyclic, so G is contained in the cyclic $\Phi^{-1}(\Phi(G))$ and is hence cyclic.   □

	name	elements
C_n	cyclic	$\left\{ e^{\frac{2a\pi i}{n}} \right\}_{a=1,\ldots,n}$
D^*_{4m}	binary dihedral	$\left\{ e^{\frac{a\pi i}{m}},\ e^{\frac{a\pi i}{m}} j \right\}_{a=1,\ldots,2m}$
T^*_{24}	binary tetrahedral	$\left\{ \pm 1, \pm i, \pm j, \pm k, \frac{1}{2}\left(\pm 1 \pm i \pm j \pm k \right) \right\}$
O^*_{48}	binary octahedral	$T^*_{24} \cup \left\{ \frac{\pm 1 \pm i}{\sqrt{2}},\ \frac{\pm 1 \pm j}{\sqrt{2}}, \ldots, \frac{\pm i \pm k}{\sqrt{2}},\ \frac{\pm j \pm k}{\sqrt{2}} \right\}$
I^*_{120}	binary icosahedral	$T^*_{24} \cup \left\{ \pm \frac{1}{2} i \pm \frac{\sqrt{5}-1}{4} j \pm \frac{\sqrt{5}+1}{4} k, \ldots \right\}$

Table 12.2. Every finite subgroup of S^3 is conjugate to one of these groups. The group O^*_{48} consists of T^*_{24} and the $6 \times 4 = 24$ numbers obtained from $\frac{\pm 1 \pm i}{\sqrt{2}}$ by permuting the elements $1, i, j, k$ and varying the signs. The group I^*_{120} consists of T^*_{24} and the $12 \times 8 = 96$ numbers obtained from $\pm \frac{1}{2} i \pm \frac{\sqrt{5}-1}{4} j \pm \frac{\sqrt{5}+1}{4} k$ by permuting the elements $1, i, j, k$ with an *even* permutation and varying the signs.

The smallest non-abelian group in the list is the *quaternion group* $Q_8 = D^*_8$ consisting of the *Lipschitz units*

$$Q_8 = \{\pm 1, \pm i, \pm j, \pm k\}.$$

The quaternion group is contained in T^*_{24} as a normal subgroup, giving an exact sequence

$$(13) \qquad\qquad 0 \longrightarrow Q_8 \longrightarrow T^*_{24} \longrightarrow \mathbb{Z}_3 \longrightarrow 0.$$

The binary dihedral group D^*_{4m} contains the index-two cyclic group C_{2m}, giving an exact sequence

$$(14) \qquad\qquad 0 \longrightarrow C_{2m} \longrightarrow D^*_{4m} \longrightarrow \mathbb{Z}_2 \longrightarrow 0.$$

Remark 12.2.10. Every group Γ in Table 12.2 acts freely and isometrically on S^3 by right multiplication, hence $S^3/_\Gamma$ is an elliptic manifold with fundamental group Γ. The only perfect group in the list (*i.e.* with trivial abelianisation) is I^*_{120} and we will soon see that $S^3/_{I^*_{120}}$ is the ubiquitous Poincaré homology sphere $\Sigma(2, 3, 5)$, defined in Section 10.3.5.

12.2.3. Regular polytopes. There are six regular polytopes in dimension four, listed in Table 12.3, and the finite groups of S^3 can be used to describe five of them. The convex hull of Q_8 in $\mathbb{R}^4$ is the *16-cell*, whose dual is the *hypercube*. The group T^*_{24} consists of the vertices of the 16-cell and of the dual hypercube altogether.

The convex hull of T^*_{24} is the *24-cell*: this is the unique self-dual regular polytope in all dimensions $n \geqslant 3$ which is not a simplex! See Figure 12.1. The group O^*_{48} consists of the vertices of the 24-cell and of its dual. Finally, the convex hull of I^*_{120} is the *600-cell*, whose dual is the *120-cell*.

name	vertices	edges	faces	facets	Schläfli
simplex	5	10	10	5 tetrahedra	$\{3, 3, 3\}$
hypercube	16	32	24	8 cubes	$\{4, 3, 3\}$
16-cell	8	24	32	16 tetrahedra	$\{3, 3, 4\}$
24-cell	24	96	96	24 octahedra	$\{3, 4, 3\}$
120-cell	600	1200	720	120 dodecahedra	$\{5, 3, 3\}$
600-cell	120	720	1200	600 tetrahedra	$\{3, 3, 5\}$

Table 12.3. The six regular polytopes in dimension four, with their Schläfli notation (see Section 3.3.3). The groups Q_8, T_{24}^*, and I_{120}^* consist of the vertices of the 16-cell, 24-cell, and 600-cell respectively.
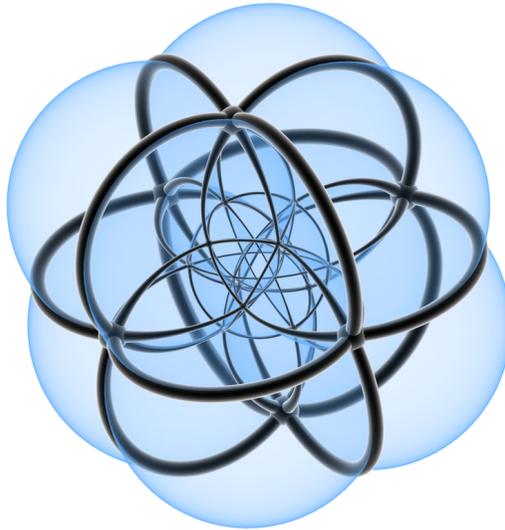
Figure 12.1. A stereographic projection of the tessellation of S^3 induced by the 24-cell. Its 24 vertices form the binary tetrahedral group T_{24}^*. Its facets are 24 regular octahedra.

12.2.4. Classification of elliptic 3-manifolds. We now want to classify all the elliptic 3-manifolds. We start with a linear algebra exercise.

Exercise 12.2.11. Let $A \in O(4)$. Then

- if $\det A = -1$, it fixes pointwise a line,

- if $\det A = +1$, it preserves two orthogonal planes and acts as a rotation on each.

Every elliptic 3-manifold is isometric to S^3/Γ for some finite subgroup $\Gamma < O(4)$ acting freely. The exercise implies that $\Gamma < SO(4)$, since orientation-reversing elements fix some point in S^3. Therefore we get:

Corollary 12.2.12. *Every elliptic 3-manifold is orientable.*

We have already encountered the lens space $L(p, q) = S^3/\Gamma$ where the cyclic group $\Gamma = C_{p,q}$ is generated by the isometry

$$(z, w) \longmapsto (\omega z, \omega^q w) \quad \text{with } \omega = e^{\frac{2\pi i}{p}}.$$

Another consequence of the exercise is the following.

Corollary 12.2.13. *Every finite abelian group $\Gamma < SO(4)$ acting freely is conjugate to $C_{p,q}$ for some coprime p, q.*

Proof. An abelian Γ fixes two orthogonal planes U and V and acts on each as rotations. The restriction map $\Gamma \to \text{Isom}^+(U)$ is injective because Γ acts freely on S^3. Therefore Γ is cyclic, generated by an element that acts on both planes by rotations of order $p = |\Gamma|$, so Γ is conjugate to $C_{p,q}$. $\square$

We now want to construct non-abelian examples. It is convenient to look at $SO(4)$ as the image of $S^3 \times S^3$ via the surjective map Ψ defined above. Which elements of $S^3 \times S^3$ act freely on S^3? The answer is particularly simple.

Proposition 12.2.14. *The isometry $\Psi(q_1, q_2)$ of S^3 has a fixed point if and only if q_1 and q_2 are conjugate in S^3.*

Proof. We have $q_1 x q_2^{-1} = x$ for some $x \in S^3$ if and only if $q_1 = x q_2 x^{-1}$, that is q_1 and q_2 are conjugate. $\square$

The following corollary will help to rule out many cases.

Corollary 12.2.15. *If q_1 and q_2 have both order four, the isometry $\Psi(q_1, q_2)$ has a fixed point.*

We now construct more examples.

Proposition 12.2.16. *If two finite subgroups $G, H < S^3$ have coprime orders, the image $\Psi(G \times H)$ acts freely on S^3.*

Proof. If $(g_1, g_2) \in G \times H$ is non-trivial, the elements g_1 and g_2 have coprime and hence distinct orders, thus they are not conjugate. $\square$

Corollary 12.2.17. *If G is a group from Table 12.2 and n is coprime with the order of G, then $\Psi(G \times C_n)$ acts freely on S^3.*

Γ	conditions	$\lvert\Gamma\rvert$	Z	$\Gamma/_Z$
$C_{p,q}$	$p > 0,\ (p,q) = 1$	p	C_p	$\{e\}$
$\Psi(D_{4m}^* \times C_n)$	$m > 1, n > 0,\ (4m, n) = 1$	$4mn$	C_{2n}	D_{2m}
$\Gamma \underset{2}{\leq} \Psi(D_{4m}^* \times C_{4n})$	$m > 1, n > 0,\ n$ even, $(m, n) = 1$	$4mn$	C_{2n}	D_{2m}
$\Psi(T_{24}^* \times C_n)$	$n > 0,\ (24, n) = 1$	$24n$	C_{2n}	T_{12}
$\Gamma \underset{3}{\leq} \Psi(T_{24}^* \times C_{6n})$	$n > 0,\ n$ odd, $3\mid n$	$24n$	C_{2n}	T_{12}
$\Psi(O_{48}^* \times C_n)$	$n > 0,\ (48, n) = 1$	$48n$	C_{2n}	O_{24}
$\Psi(I_{120}^* \times C_n)$	$n > 0,\ (120, n) = 1$	$120n$	C_{2n}	I_{60}

Table 12.4. The finite subgroups $\Gamma < \mathrm{SO}(4)$ that act freely on S^3, up to conjugation in $\mathrm{O}(4)$. For each we show its centre Z and the quotient $\Gamma/_Z$. If Γ is not cyclic, the two groups Z and $\Gamma/_Z$ determine Γ up to conjugation (note that n satisfies some congruence equality which separates lines 2,3 and 4,5). The symbol $\Gamma <_i G$ indicates that Γ has index i.

This corollary produces most of the non-cyclic subgroups of $\mathrm{SO}(4)$ acting freely on S^3, but not all of them! Table 12.4 lists all the finite subgroups of $\mathrm{SO}(4)$ acting freely, up to conjugation in $\mathrm{O}(4)$. The first line shows the cyclic groups $C_{p,q}$, and the other families of groups are all of type $\Psi(G \times C_n)$, except the third and fifth line which we now explain.

Remark 12.2.18. For all groups in Table 12.4 of type $G \times C_n$, we have $\Psi(G \times C_n) \cong G \times C_n$ because n is odd and hence $(-1, -1) \notin G \times C_n$.

In the third line of Table 12.4, the group Γ is an index-two subgroup of $\Psi(D_{4m}^* \times C_{4n})$. It is the image along Ψ of the kernel of the map

$$D_{4m}^* \times C_{4n} \longrightarrow \mathbb{Z}_2$$
$$(g_1, g_2) \longmapsto f_1(g_1) + f_2(g_2)$$

where f_1 is the map in (14) and f_2 is the surjective homomorphism $C_{4n} \to \mathbb{Z}_2$. Analogously, the group Γ in the fifth line is an index-three subgroup of $\Psi(T_{24}^* \times C_{6n})$: it is the image along Ψ of the kernel of the map

$$T_{24}^* \times C_{6n} \longrightarrow \mathbb{Z}_3$$
$$(g_1, g_2) \longmapsto f_1(g_1) + f_2(g_2)$$

where f_1 is the map in (13) and f_2 is any surjective homomorphism $C_{6n} \to \mathbb{Z}_3$. Both f_1 and f_2 are well-defined only up to automorphisms of $\mathbb{Z}_3$, but the kernel is well-defined up to conjugation in $S^3 \times S^3$ anyway: there are four possibilities and they are related by conjugations via the elements

$(1, j)$ and $\left(\frac{1+i}{\sqrt{2}}, 1\right)$, as one can easily check (these conjugations permute the cosets of $C_{2n} \triangleleft C_{6n}$ and $Q_8 \triangleleft T_{24}^*$).

Proposition 12.2.19. *The finite subgroups of $SO(4)$ shown in Table 12.4 act freely on S^3.*

Proof. We only need to prove this for the groups Γ that belong to the third and fifth family. Concerning the third, pick an element $(g_1, g_2) \in S^3 \times S^3$ that projects to a non-trivial element in Γ. We want to prove that g_1 and g_2 have distinct orders and are hence non-conjugate.

By hypothesis $f_1(g_1) + f_2(g_2) = 0$ in $\mathbb{Z}_2$ and hence $f_1(g_1) = f_2(g_2)$ is either 0 or 1. In the first case we get $g_1 \in C_{2m}$ and $g_2 \in C_{2n}$ which have distinct orders since $(m, n) = 1$ and $(g_1, g_2) \neq \pm(1, 1)$. In the second case $g_1 \in D_{4m}^* \setminus C_{2m}$ has order four (check from Table 12.2) and $g_2 \in C_{4n} \setminus C_{2n}$ does not have order four since n is even.

The fifth family is treated similarly. Pick $(g_1, g_2) \in \Gamma$ with non-trivial image in $SO(4)$. If $f_1(g_1) = f_2(g_2) = 0$ then $g_1 \in Q_8$ has order 1, 2, or 4 and $g_2 \in C_{2n}$ does not have order 4 since n is odd. If $f_1(g_1) = -f_2(g_2) \neq 0$ then $g_1 \in T_{24}^* \setminus Q_8$ has order 3 or 6 (check from Table 12.2) while $g_2 \in C_{6n} \setminus C_{2n}$ cannot have order 3 or 6 since 3 divides n. $\square$

We now show that Table 12.4 exhausts all possibilities.

Proposition 12.2.20. *Every finite subgroup of $SO(4)$ acting freely on S^3 is conjugate in $O(4)$ to one in Table 12.4.*

Proof. Let a finite $\Gamma < SO(4)$ act freely on S^3. We consider its counterimage $G = \Psi^{-1}(\Gamma)$ in $S^3 \times S^3$.

We note that the orientation-reversing isometry $q \mapsto q^{-1}$ of S^3 conjugates $\Psi(q_1, q_2)$ to $\Psi(q_2, q_1)$, since

$$x \longmapsto (q_1 x^{-1} q_2^{-1})^{-1} = q_2 x q_1^{-1}.$$

Therefore via conjugation in $O(4)$ we may swap the factors of $S^3 \times S^3$.

If G is a product $G = G_1 \times G_2$, then G_1 and G_2 are some groups from Table 12.2. If they are both cyclic, then Γ is abelian and we conclude by Corollary 12.2.13. If they are both non-cyclic, they both contain order-4 elements and we get a contradiction from Proposition 12.2.15.

Therefore $G = G_1 \times C_k$ with G_1 equal to D_{4m}^*, T_{24}^*, O_{48}^*, or I_{120}^*, and k not divisible by four. Since $(-1, -1) \in G$ we get $k = 2n$, and n odd implies that $\Psi(G_1 \times C_n) = \Psi(G_1 \times C_{2n})$. Moreover n is coprime with the order of G_1 otherwise Proposition 12.2.14 would easily give a fixed point. Therefore we get a product group as in Table 12.4.

We are left to consider the case G is not a product. Let $G_1 \times G_2$ (resp. $G_1' \times G_2'$) be the smallest (resp. biggest) product subgroup of $S^3 \times S^3$ such that

$$G_1' \times G_2' < G < G_1 \times G_2.$$

The subgroup G_1 (resp. G_1') consists of all $g_1 \in S^3$ such that $(g_1, g_2) \in G$ for some g_2 (resp. such that $(g_1, 1) \in G$). It is easy to check that $G_i' \triangleleft G_i$ and

$$G/_{G_1' \times G_2'} \cong G_1/_{G_1'} \cong G_2/_{G_2'}.$$

Each G_i and G_i' is finite and hence conjugate to one in Table 12.2. If both G_1 and G_2 are cyclic, the group Γ is abelian and we are done. We henceforth suppose that G_1 and G_2 are not both cyclic.

By the same reasoning above, up to reordering G_2' contains no order-4 elements and hence $G_2' = C_s$ with s not divisible by four.

We now prove that G_1' contains all the order-4 elements of G_1. To do so we pick an order-4 element $g_1 \in G_1$ and prove that $\pm g_1 \in G_1'$: this suffices since $g_1 = (-g_1)^3$. We have $(g_1, g_2) \in G$ for some g_2 of some order $2^t k$ with k odd. We have $(g_1^k, g_2^k) = (\pm g_1, g_2^k)$, so up to replacing g_2 with g_2^k we may suppose $k = 1$ and g_2 has order 2^t.

We have $(1, g_2^4) = (g_1^4, g_2^4) \in G$ and hence $g_2^4 \in G_2'$. Suppose s is odd. Therefore $g_2^4 \in G_2' = C_s$ is trivial and g_2 has order 2^t; thus g_2 has order $1, 2$, or 4. It cannot have order 4 by Corollary 12.2.15, therefore $g_2 = \pm 1$ and hence $\pm g_1 \in G_1'$ as required.

If s is even we get the same conclusion: now $(-1, g_2^2) = (g_1^2, g_2^2) \in G$ hence $(1, -g_2^2) \in G$ gives $-g_2^2 \in G_2'$ and thus $g_2^2 \in G_2'$ since s is even. Again this implies that g_2 has order $1, 2$, or 4 and we conclude as above.

We have proved that G_1' contains all order-4 elements of G_1. The groups D_{4n}^*, O_{48}^*, and I_{120}^* are generated by their order-4 elements (exercise), hence $G_1' \neq G_1$ implies that G_1 is not one of them. Then it is either cyclic or T_{24}^*.

If $G_1 = C_h$ is cyclic, then G_1' and $G_1/_{G_1'} \cong G_2/_{G_2'}$ also are. By assumption G_2 is not cyclic, but it contains the cyclic G_2' with non-trivial cyclic quotient: the only possibility from Table 12.2 is that $G_2 = D_{4m}^*$ and $G_2' = C_{2m}$. Therefore G is an index-two subgroup of $G_1 \times G_2 = C_h \times D_{4m}^*$ as in the third line of Table 12.4.

We need to prove that $h = 4n$ with n even and $(n, m) = 1$. Since $G_1' < G_1$ has index two, h is even. Since $(1, -1) \in G_1' \times C_{2m} < G$, we also have $(-1, 1) \in G$ and hence $-1 \in G_1'$. Therefore $|G_1'|$ is even and $4|h$. Since G_1' contains all order-4 elements of G_1, four divides $|G_1'|$ and hence $8|h$. Therefore $h = 4n$ with n even, as required. Moreover $(m, n) = 1$ since $G_1' \times G_2' = C_{2n} \times C_{2m} \subset G$.

If $G_1 = T_{24}^*$ the order-4 elements generate the index-three subgroup Q_8 and hence $G_1' = Q_8$ and $G_2/_{G_2'} \cong G_1/_{G_1'} \cong \mathbb{Z}_3$. Recall that $G_2' = C_s$ with s not divisible by four. As above, $-1 \in G_1'$ implies $-1 \in G_2'$ and hence $s = 2n$ with n odd. Therefore G_2 has order $6n$ and must hence be cyclic.

The group G is an index-3 subgroup of $G_1 \times G_2 = T_{24}^* \times C_{6n}$ as in the fifth line of Table 12.4. We must have $3|n$ otherwise G contains an element (g_1, g_2) with both g_1, g_2 having order 3 (see the proof of Proposition 12.2.19). □

We can finally summarise our discoveries:

Corollary 12.2.21. *Table 12.4 lists all the finite subgroups $\Gamma < SO(4)$ acting freely on S^3 up to conjugation in $O(4)$, without repetitions.*

Proof. There are no repetitions because the non-cyclic groups listed are all non-isomorphic: the isomorphism types of the centre Z and Γ/Z suffice to determine Γ in that list, see Table 12.4 (note that n satisfies some congruence equality which separates the lines 2,3 and 4,5).

To compute $|\Gamma|$, we use Remark 12.2.18 and note that $(-1, -1)$ belongs to $(D_{4m}^* \times C_{4n})$ and $(T_{24}^* \times C_{6n})$.

To compute Z and Γ/Z, we note that in all the non-cyclic cases Γ is the image of a subgroup in some product $G^* \times C_{kn}$ which is "diagonal", in the sense that it maps surjectively to both factors G^* and C_{kn}. The map onto G^* pushes-forward to a surjection $\Gamma \to G$, whose kernel is easily detected as being the centre Z and isomorphic to C_{2n} in all cases. $\square$

12.2.5. Seifert fibrations of elliptic 3-manifolds. We now turn back to Seifert manifolds and prove Theorem 12.2.1. Let $S^1 \subset S^3$ be the unit complex numbers. Recall that $S^3 \times S^3$ acts on S^3 via Ψ. The following proposition is crucial: it says that a big Lie subgroup of $S^3 \times S^3$ preserves the Hopf fibration; later on we will discover that every finite subgroup Γ of $S^3 \times S^3$ acting freely may be conjugated into this big group and hence every quotient S^3/Γ inherits a Seifert structure from the Hopf fibration, hence every elliptic three-manifold is Seifert.

Proposition 12.2.22. *The group $S^1 \times S^3$ preserves the Hopf fibration.*

Proof. Represent quaternions as pairs (z_1, z_2) of complex numbers. The Hopf fibration $S^3 \to \mathbb{CP}^1$ is the map $(z_1, z_2) \to [z_1, z_2]$. One checks easily that quaternion multiplication acts as follows:

$$(z_1, z_2)(w_1, w_2) = (z_1 w_1 - z_2 \bar{w}_2, z_2 \bar{w}_1 + z_1 w_2).$$

Therefore right-multiplication by $(w_1, w_2) \in S^3$ acts $\mathbb{C}$-linearly on $\mathbb{C}^2$ and hence preserves the Hopf fibration.

Left multiplication by (w_1, w_2) is *not* $\mathbb{C}$-linear in general, but it is so when $(w_1, w_2) = (w_1, 0) \in S^1$, since $(w_1, 0)(z_1, z_2) = (w_1 z_1, w_1 z_2)$. Therefore $S^1 \times S^3$ preserves each fibre of the Hopf fibration. $\square$

We can finally prove Theorem 12.2.1.

Theorem 12.2.23. *A closed 3-manifold M admits an elliptic metric if and only if it is a Seifert manifold with $e \neq 0$ and $\chi > 0$.*

Proof. If M is elliptic then $M = S^3/\Gamma$ for some finite subgroup $\Gamma < SO(4)$ acting freely. If Γ is conjugate to $C_{p,q}$ we get a lens space and we are done. Otherwise Table 12.4 shows that Γ is conjugate to the image of a subgroup of $S^1 \times S^3$ (because $C_n < S^1$). Therefore up to conjugation Γ

fibration	condition	Γ
$\big(S^2, (2,1), (2,1), (p,q)\big)$	$p + q > 0$ odd	$\Psi\big(D_{4p}^* \times C_{p+q}\big)$
$\big(S^2, (2,1), (2,1), (p,q)\big)$	$p + q > 0$ even	$\Gamma \underset{2}{<} \Psi\big(D_{4p}^* \times C_{4(p+q)}\big)$
$\big(S^2, (2,1), (3,1), (3,q)\big)$	$3 \nmid 5 + 2q$	$\Psi\big(T_{24}^* \times C_{|5+2q|}\big)$
$\big(S^2, (2,1), (3,1), (3,q)\big)$	$3 \mid 5 + 2q > 0$	$\Gamma \underset{3}{<} \Psi\big(T_{24}^* \times C_{6(5+2q)}\big)$
$\big(S^2, (2,1), (3,1), (4,q)\big)$		$\Psi\big(O_{48}^* \times C_{|10+3q|}\big)$
$\big(S^2, (2,1), (3,1), (5,q)\big)$		$\Psi\big(I_{120}^* \times C_{|25+6q|}\big)$

Table 12.5. The non-lens elliptic manifolds, listed without repetitions. For each we show its Seifert fibration and fundamental group $\Gamma < SO(4)$. Here $p > 1$. The integer q may be negative, if not forbidden explicitly. In the fourth line 3 divides $5 + 2q$, which must be positive.

preserves the Hopf fibration of S^3, which descends to a Seifert fibration. We have $\chi > 0$ and $e \neq 0$ because the universal cover of M is S^3.

Conversely, we now show that every Seifert manifold with $\chi > 0$ and $e \neq 0$ arises as an elliptic manifold $S^3/_\Gamma$. The Seifert manifolds were listed in Table 10.6, and the corresponding Γ is shown in Table 12.5. To verify the correspondence, note that the centre Z of $\Gamma = \pi_1(M)$ and the quotient $\Gamma/_Z$ are shown in Table 12.4 and they fully determine Γ. Proposition 10.4.18 says that $Z < \pi_1(M)$ is generated by a regular fibre and $\Gamma/_Z$ is the orbifold fundamental group of the base. $\qquad\square$

In particular we have the following:

$$\pi_1\big(S^2, (2,1), (2,1), (2,-1)\big) = Q_8,$$
$$\pi_1\big(S^2, (2,1), (2,1), (p,1-p)\big) = D_{4p}^*,$$
$$\pi_1\big(S^2, (2,1), (3,1), (3,-2)\big) = T_{24}^*,$$
$$\pi_1\big(S^2, (2,1), (3,1), (4,-3)\big) = O_{48}^*,$$
$$\pi_1\big(S^2, (2,1), (3,1), (5,-4)\big) = I_{120}^*.$$

The latter is Poincaré's homology sphere, the unique elliptic three-manifold with perfect fundamental group (see Section 10.3.5).

12.3. Flat three-manifolds

We turn to flat three-manifolds. In dimension two, every orientation-preserving isometry of $\mathbb{R}^2$ is a translation, and this easily implies that every flat orientable closed surface is a torus. In dimension three we also have *rototranslations*, which produce more orientable manifolds.

We now classify all the closed flat orientable three-manifolds up to diffeomorphism and prove the following.

Theorem 12.3.1. *A closed orientable 3-manifold M admits a flat metric if and only if it is a Seifert manifold with $e = \chi = 0$.*

There are six such manifolds, listed in Table 10.3.

12.3.1. Classification. Every closed flat 3-manifold is isometric to $\mathbb{R}^3/_\Gamma$ for some crystallographic group $\Gamma < \mathrm{Isom}(\mathbb{R}^3)$ acting freely, see Section 4.4.4.

Exercise 12.3.2. Every element in Γ is either a translation or a roto-translation (defined in Example 4.4.7).

We prove one half of Theorem 12.3.1.

Proposition 12.3.3. *Every closed orientable flat 3-manifold is a Seifert manifold with $e = \chi = 0$.*

Proof. We have $M = \mathbb{R}^3/_\Gamma$. Recall the exact sequence

$$0 \longrightarrow H \longrightarrow \Gamma \longrightarrow r(\Gamma) \longrightarrow 0$$

where $H \triangleleft \Gamma$ is the translation subgroup and $r(\Gamma) < SO(3)$ is finite by Proposition 4.4.9. We now prove that Γ preserves a foliation of $\mathbb{R}^3$ into parallel lines that projects to a Seifert structure on M.

If $r(\Gamma)$ is trivial, then $\Gamma = H$ consists of translations and preserves many foliations into parallel lines that project to a Seifert structure on the quotient 3-torus M. If $r(\Gamma)$ is non-trivial, it is isomorphic to C_n, D_{2m}, T_{12}, O_{24}, or I_{60}. If $r(\Gamma) = C_n$ or D_{2m}, it has a common fixed vector line $l \subset \mathbb{R}^3$ and Γ preserves the foliation of lines parallel to l.

If $r(\Gamma) = T_{12}, O_{24}$, or I_{60} we obtain a contradiction as follows. In all cases we have $T_{12} \subset r(\Gamma)$. The group T_{12} consists of the identity, the π-rotations along the three coordinate axis, and the $\pm\frac{2\pi}{3}$-rotations along the axis spanned by $(1, 1, 1), (1, -1, -1), (-1, 1, -1)$, and $(1, -1, -1)$.

Pick a rototranslation $h \in \Gamma$ with axis l parallel to $(1, 1, 1)$, and up to conjugating Γ by a translation we may suppose that l contains the origin $0 \in \mathbb{R}^3$. We have $h(0) = (d, d, d)$ for some $d \neq 0$. Since h^3 is a translation we get $(3d, 3d, 3d) \in H$.

The group Γ and hence T_{12} acts on H via conjugation. Therefore H is T_{12}-symmetric and $(3d, -3d, -3d) \in H$ using the π-rotation along the first axis. Hence $t = (6d, 0, 0) = (3d, 3d, 3d) + (3d, -3d, -3d) \in H$. The composition $t \circ h^2$ has a fixed point, because it sends 0 to $(2d, 2d, 2d) - (6d, 0, 0) = (-4d, 2d, 2d)$ which is orthogonal to l: a contradiction.

In all cases M has a Seifert structure. By Bieberbach's Theorem (stated as Corollary 4.4.11) the manifold M is finitely covered by the 3-torus, and hence $\chi = e = 0$ by Proposition 10.3.26. $\square$

We now prove the other half.
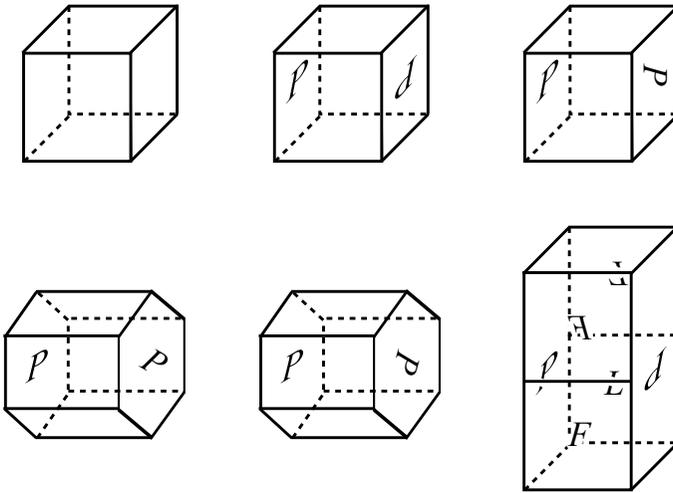
Figure 12.2. The six closed orientable flat 3-manifolds, up
to diffeomorphism. Each is constructed by pairing isometri-
cally the faces of a polyhedron in $\mathbb{R}^3$ according to the labels.
When a face has no label, it is simply paired to its oppo-
site by a translation. The polyhedra shown here are three
cubes, two prisms with regular hexagonal basis, and one par-
allelepiped made of two cubes.

Proposition 12.3.4. *Every closed Seifert manifold M with $e = \chi = 0$
admits a flat metric.*

Proof. There are six such Seifert manifolds up to diffeomorphism,
listed in Table 10.3. We build a flat metric for each in Figure 12.2.

The figure shows six flat manifolds, constructed by identifying isomet-
rically the faces of a polyhedron in $\mathbb{R}^3$. The reader is invited to check that
each construction gives indeed a flat manifold, by verifying that the flat
metric extends to the edges and to the vertices.

In all cases the foliation by parallel horizontal lines (orthogonal to the
P faces) descends to a Seifert fibration on the flat manifold. By looking at
these lines one checks that the base surface of the fibration is respectively

$$T, \ (S^2, 2, 2, 2, 2), \ (S^2, 2, 4, 4), \ (S^2, 2, 3, 6), \ (S^2, 3, 3, 3), \ (\mathbb{RP}^2, 2, 2).$$

These orbifolds are obtained respectively from the figures by considering:
the square torus, its quotient via a π-rotation, via a $\frac{\pi}{2}$-rotation, the quo-
tient of a hexagon torus by a $\frac{\pi}{3}$-rotation, by a $\frac{2\pi}{3}$-rotation, and the quotient
of a Klein bottle via a π-rotation.

These flat Seifert manifolds have $e = 0$ by Proposition 12.3.3, and
hence they are precisely those listed in Table 10.3.   □
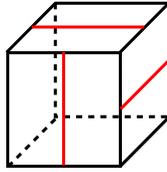
Figure 12.3. The Hantsche-Wendt manifold $M = \mathbb{R}^3/_\Gamma$. The group Γ is generated by three rototranslations along the three red axis shown, each of angle π and with unit displacement (this is a unit cube). A fundamental domain for Γ is made of two unit cubes as in Figure 12.2-(bottom, right).

We have proved Theorem 12.3.1.

Remark 12.3.5. The rotational image $r(\Gamma) < SO(3)$ of Γ for the six flat manifolds $M = \mathbb{R}^3/_\Gamma$ constructed in Figure 12.2 is respectively $\{e\}$, C_2, C_4, C_6, C_3, and the dihedral $D_4 = C_2 \times C_2$. The sixth manifold is called the *Hantzsche-Wendt manifold*: the group Γ is generated by three rototranslations as in Figure 12.3.

In contrast with the hyperbolic case, a finite-volume complete flat manifold is necessarily closed (see Proposition 4.4.12). The Seifert manifolds with boundary and $\chi = 0$ are diffeomorphic to the bundles $T \times I$ and $K \,\widetilde{\times}\, I$. Their interiors $T \times \mathbb{R}$ and $K \,\widetilde{\times}\, \mathbb{R}$ may be given an *infinite*-volume complete flat structure. The lack of a finite-volume complete flat structure for $K \,\widetilde{\times}\, \mathbb{R}$ is a reason for preferring the geometric decomposition to the canonical torus decomposition, see Section 11.5.3.

12.4. The product geometries

The eight three-dimensional geometries include the three isotropic ones, plus five more. Among the five non-isotropic geometries, two are products of lower-dimensional geometries. We analyse them here.

12.4.1. $S^2 \times \mathbb{R}$ **geometry.** We equip $S^2 \times \mathbb{R}$ with the product metric. The product $S^2 \times \mathbb{R}$ is the poorest of the eight geometries, in the sense that there are very few manifolds modelled on $S^2 \times \mathbb{R}$.

Recall that the sectional curvature is a number assigned to every plane in the tangent space of every point $p \in S^2 \times \mathbb{R}$. One such plane is *horizontal* if it is tangent to the S^2 factor and *vertical* if it contains the line tangent to the $\mathbb{R}$ factor.

Proposition 12.4.1. *The sectional curvatures of horizontal and vertical planes are* 1 *and* 0, *respectively.*

Proof. Let $\gamma \subset S^2$ be a closed geodesic. The surfaces $S^2 \times y$ and $\gamma \times \mathbb{R}$ are totally geodesic, because they are fixed by some isometric reflections

of $S^2 \times \mathbb{R}$. Thus the sectional curvatures of horizontal and vertical planes equal the gaussian curvatures of these surfaces, which are 1 and 0.   □

Proposition 12.4.2. *We have*

$$\mathsf{Isom}(S^2 \times \mathbb{R}) = \mathsf{Isom}(S^2) \times \mathsf{Isom}(\mathbb{R}).$$

Proof. We certainly have the inclusion $\supset$, which gives to every point $x \in S^2 \times \mathbb{R}$ a stabiliser in $\mathsf{Isom}^+(S^2 \times \mathbb{R})$ isomorphic to $\mathrm{SO}(2) \rtimes C_2$, a proper maximal subgroup of $\mathrm{SO}(3)$ by Proposition 6.2.15.

If there were more isometries than that, there would be more fixing p since they act transitively on $S^2 \times \mathbb{R}$ and the stabiliser would be the whole of $\mathrm{SO}(3)$, a contradiction because the sectional curvature of $S^2 \times \mathbb{R}$ is not constant.   □

Since the isometry groups of S^2 and $\mathbb{R}$ have two connected components each, the group $\mathsf{Isom}(S^2 \times \mathbb{R})$ has four connected components, two of which are orientation-preserving.

Proposition 12.4.3. *An orientable manifold M admits a finite-volume $S^2 \times \mathbb{R}$ geometry $\Longleftrightarrow M$ is a closed Seifert manifold with $e = 0$ and $\chi > 0$.*

Proof. The closed Seifert manifolds with $e = 0$ and $\chi > 0$ are just $S^2 \times S^1$ and $\mathbb{RP}^2 \mathbin{\widetilde{\times}} S^1$, and they are diffeomorphic to $(S^2 \times \mathbb{R})/_\Gamma$ where Γ is generated respectively by

$$\big\{(\mathrm{id}, \tau)\big\}, \qquad \big\{(\iota, r), (\iota, r')\big\}$$

where τ is any translation, ι is the antipodal map and r, r' are reflections with respect to distinct points in $\mathbb{R}$.

Conversely, pick an orientable $M = (S^2 \times S^1)/_\Gamma$. The discrete subgroup $\Gamma < \mathsf{Isom}(S^2) \times \mathsf{Isom}(\mathbb{R})$ preserves the foliation into spheres $S^2 \times \{x\}$ which descends into a foliation into spheres and/or projective planes for M. Therefore M decomposes into orientable interval bundles $S^2 \times I$ and $\mathbb{RP}^2 \mathbin{\widetilde{\times}} I$, and is hence either $S^2 \times S^1$ or $\mathbb{RP}^2 \mathbin{\widetilde{\times}} S^1$.   □

12.4.2. $\mathbb{H}^2 \times \mathbb{R}$ geometry. We give $\mathbb{H}^2 \times \mathbb{R}$ the product metric. The discussion of the previous section applies as is to this case, showing that horizontal and vertical planes in the tangent spaces have sectional curvature -1 and 0. This in turn implies that

$$\mathsf{Isom}(\mathbb{H}^2 \times \mathbb{R}) = \mathsf{Isom}(\mathbb{H}^2) \times \mathsf{Isom}(\mathbb{R})$$

has four connected components, two being orientation-preserving. It is convenient to write the exact sequence

$$0 \longrightarrow \mathsf{Isom}(\mathbb{R}) \longrightarrow \mathsf{Isom}(\mathbb{H}^2 \times \mathbb{R}) \xrightarrow{\ p\ } \mathsf{Isom}(\mathbb{H}^2) \longrightarrow 0.$$

A discrete group $\Gamma < \mathsf{Isom}(X)$ is *cofinite* if $X/_\Gamma$ has finite volume.

Proposition 12.4.4. *A discrete group $\Gamma < \mathsf{Isom}(\mathbb{H}^2 \times \mathbb{R})$ is cofinite if and only if both $p(\Gamma)$ and $\Gamma \cap \ker p$ are discrete and cofinite.*

Proof. If $p(\Gamma)$ is discrete we get

(15) $\quad \mathrm{Vol}\big((\mathbb{H}^2 \times \mathbb{R})/_\Gamma\big) = \mathrm{Area}(\mathbb{H}^2/_{p(\Gamma)}) \times \mathrm{Length}\big(\mathbb{R}/_{\Gamma \cap \ker p}\big).$

This surprisingly simple formula is proved by picking a fundamental domain $D \subset \mathbb{H}^2$ for $p(\Gamma)$ and noticing that

$$\mathrm{Vol}\big((\mathbb{H}^2 \times \mathbb{R})/_\Gamma\big) = \mathrm{Vol}\big(p^{-1}(D)/_{\Gamma \cap \ker p}\big).$$

We deduce that Γ is cofinite if and only if both $p(\Gamma)$ and $\Gamma \cap \ker p$ are.

If $p(\Gamma)$ is not discrete, we prove that Γ cannot be cofinite. Up to replacing Γ with an index-four subgroup we may suppose

$$\Gamma < \mathrm{Isom}^+(\mathbb{H}^2) \times \mathrm{Isom}^+(\mathbb{R}) = \mathrm{Isom}^+(\mathbb{H}^2) \times \mathbb{R}.$$

Pick a neighbourhood $U \subset \mathrm{Isom}^+(\mathbb{H}^2) \times \mathbb{R}$ of e such that $[U, U] \subset U$ and $U \cap \Gamma = \{e\}$. Let $f, g \in \Gamma$ be two elements. We now prove that if $p(f), p(g) \in p(U)$ then f and g commute. We note that the commutator $[f, g]$ depends only on the images $p(f)$ and $p(g)$, and since they lie in $p(U)$ we may suppose that $f, g \in U$ and get $[f, g] \in U \cap \Gamma$, which must be trivial.

The elements in $p(\Gamma) \cap p(U)$ commute. Two non-trivial elements in $\mathrm{Isom}^+(\mathbb{H}^2)$ commute if and only if they are both hyperbolic, parabolic, or elliptic fixing the same line, point in $\partial \mathbb{H}^2$, or point in $\mathbb{H}^2$. Therefore all the isometries in $p(\Gamma) \cap p(U)$ are of the same type and fix the same line or point.

Analogously, for every $f \in \Gamma$ we pick a neighbourhood U_f of e such that $[f, U_f] \subset U$ and conclude that f commutes with all the elements in Γ projecting to $p(U_f)$. (There are non-trivial such elements since $p(\Gamma)$ is not discrete.) Therefore $p(f)$ also fixes the same line or point as above.

We have proved that $p(\Gamma)$ fixes a line, a horocycle, or a point in $\mathbb{H}^2$ and hence fixes its inverse image in $\mathbb{H}^2 \times \mathbb{R}$ which is a line or a Euclidean plane. Moreover Γ acts freely and proper discontinuously on it: hence $\Gamma = \mathbb{Z}$ or $\mathbb{Z}^2$ up to finite index and it is easy to deduce that Γ is not cofinite. $\qquad\square$

Corollary 12.4.5. *If the interior of a compact orientable manifold M admits a finite-volume complete $\mathbb{H}^2 \times \mathbb{R}$ geometry then M is a Seifert manifold with $\chi < 0$. If M is closed then also $e = 0$.*

Proof. We have $\mathrm{int}(M) = (\mathbb{H}^2 \times \mathbb{R})/_\Gamma$ with Γ cofinite: by Proposition 12.4.4 the group $\Gamma \cap \ker p$ quotients every line $\{x\} \times \mathbb{R}$ to a circle in M, giving a Seifert fibration $M \to S$ onto the finite-area orbifold $S = \mathbb{H}^2/_{p(\Gamma)}$. We have $\chi(S) < 0$, and either $e(M) = 0$ or $\partial M \neq \varnothing$ because $\mathbb{H}^2 \times y$ projects to a section for $M \to S$. $\qquad\square$

We now prove the converse of Corollary 12.4.5.

Proposition 12.4.6. *If M is a Seifert manifold with $\chi < 0$ and either $\partial M \neq \varnothing$ or $e = 0$, the interior of M admits a finite-volume complete $\mathbb{H}^2 \times S^1$ geometry.*

Proof. By hypothesis there is a section Σ of $M \to S$, which is the fibre of a bundle $M \to O$ over a 1-orbifold O, see Section 11.4.4. The two structures give two exact sequences

$$0 \longrightarrow K \longrightarrow \pi_1(M) \xrightarrow{f} \pi_1(S) \longrightarrow 0,$$

$$0 \longrightarrow \pi_1(\Sigma) \longrightarrow \pi_1(M) \xrightarrow{g} \pi_1(O) \longrightarrow 0.$$

Since $\chi(S) < 0$ we may write $S = \mathbb{H}^2/\Gamma$ and identify $\pi_1(S)$ with $\Gamma <$ Isom($\mathbb{H}^2$). Analogously we consider $\pi_1(O)$ inside Isom($\mathbb{R}$). The map

$$(f, g) \colon \pi_1(M) \longrightarrow \text{Isom}(\mathbb{H}^2) \times \text{Isom}(\mathbb{R})$$

is injective and its image is discrete and acts freely on $\mathbb{H}^2 \times \mathbb{R}$, inducing a finite-volume $\mathbb{H}^2 \times \mathbb{R}$ structure on M. $\square$

12.5. Nil geometry

There are still three non-isotropic geometries to analyse. These geometries are not products, but they have a reasonable bundle structure, so that many of the arguments of the previous section can be extended with not much effort.

We start with the geometry Nil, which is a $\mathbb{R}$-bundle over $\mathbb{R}^2$. This geometry is fully carried by a Lie group called the *Heisenberg group*.

12.5.1. The Heisenberg group. The *Heisenberg group* consists of all matrices

$$\begin{pmatrix} 1 & x & z \\ 0 & 1 & y \\ 0 & 0 & 1 \end{pmatrix}$$

with $x, y, z \in \mathbb{R}$, with the multiplication operation. This is a nilpotent (but non-abelian) Lie group (see Exercise 1.4.4) and is hence also called Nil. It is clearly diffeomorphic to $\mathbb{R}^3$. There is a Lie groups exact sequence

$$0 \longrightarrow \mathbb{R} \longrightarrow \text{Nil} \longrightarrow \mathbb{R}^2 \longrightarrow 0$$

where $\mathbb{R} = [\text{Nil}, \text{Nil}]$ is the centre of Nil and consists of all matrices with $x = y = 0$. Therefore Nil is naturally a line bundle over $\mathbb{R}^2$. We identify Nil with $\mathbb{R}^3$ using the coordinates (x, y, z). The product operation becomes

$$(x, y, z) \cdot (x', y', z') = (x + x', y + y', z + z' + xy').$$

Exercise 12.5.1. The commutator of two elements is

$$\big[(x, y, z), (x', y', z')\big] = (0, 0, xy' - x'y).$$

12.5.2. The geometry of Nil. Every left-invariant Riemannian structure on a Lie group G is constructed by fixing a scalar product at the tangent space T_eG of $e \in G$ and then extending it by left-multiplication. Here $e = (0, 0, 0)$ and we give $T_e\text{Nil} = \mathbb{R}^3$ the Euclidean scalar product. This defines a left-invariant Riemannian structure on Nil.
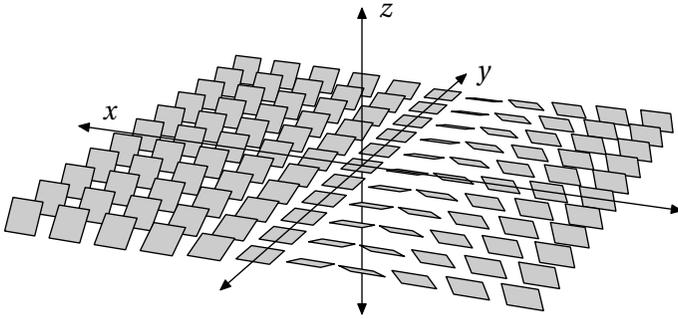
Figure 12.4. The planes orthogonal to the z axis in Nil form the standard contact structure on $\mathbb{R}^3$.

Exercise 12.5.2. The metric tensor at (x, y, z) is

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & x^2 + 1 & -x \\ 0 & -x & 1 \end{pmatrix}.$$

The following triple is an orthonormal basis at (x, y, z):

(16) $\qquad (1, 0, 0), \quad (0, 1, x), \quad (0, 0, 1).$

The planes spanned by the first two vectors form the standard *contact structure* in $\mathbb{R}^3$ shown in Figure 12.4. It is a non-integrable distribution of planes: no surface is tangent to it at every point.

We remark that the metric tensor has unit determinant at every point: therefore the volume form on Nil is the standard one on $\mathbb{R}^3$.

We can calculate the Christoffel symbols Γ_{ij}^k by hand or using a computer code; these are all zero except the following:

$$\Gamma_{22}^1 = -x, \quad \Gamma_{23}^1 = \Gamma_{32}^1 = \frac{1}{2},$$

$$\Gamma_{12}^2 = \Gamma_{21}^2 = \frac{x}{2}, \quad \Gamma_{13}^2 = \Gamma_{31}^2 = -\frac{1}{2},$$

$$\Gamma_{12}^3 = \Gamma_{21}^3 = \frac{x^2 - 1}{2}, \quad \Gamma_{13}^3 = \Gamma_{31}^3 = -\frac{x}{2}.$$

The Ricci tensor at (x, y, z) is

$$R_{ij} = \frac{1}{2} \cdot \begin{pmatrix} -1 & 0 & 0 \\ 0 & x^2 - 1 & -x \\ 0 & -x & 1 \end{pmatrix}.$$

When $x = 0$ the Ricci tensor is a diagonal matrix with entries $-\frac{1}{2}, -\frac{1}{2}, \frac{1}{2}$. For a unit vector $v \in T_p \mathrm{Nil}$, recall that $R_{ij} v^i v^j$ equals twice the average value of the sectional curvatures of the planes containing v: this average value ranges here from $-\frac{1}{4}$ to $\frac{1}{4}$ and is maximal when $v = (0, 0, \pm 1)$. This holds when $x = 0$ and hence at all points $p \in \mathrm{Nil}$ by left-multiplication.

12.5.3. The isometry group of Nil. The group Nil acts on itself iso-
metrically by left-multiplication. Left-multiplication by (a, b, c) preserves
the bundle Nil $\rightarrow \mathbb{R}^2$ and induces on $\mathbb{R}^2$ a translation by the vector (a, b).

There are also more complicate isometries of Nil that preserve the
bundle structure but induce rotations on $\mathbb{R}^2$: one such isometry φ sends
(x, y, z) to

$$\left(x \cos \theta - y \sin \theta, \ x \sin \theta + y \cos \theta, \ z + \frac{1}{2}(x^2 - y^2) \sin \theta \cos \theta - xy \sin^2 \theta\right).$$

This map preserves the bundle and induces a rotation on $\mathbb{R}^2$.

Proposition 12.5.3. *The map φ is an isometry of* Nil.

Proof. Set $(x', y', z') = \varphi(x, y, z)$. The differential $d\varphi$ at (x, y, z) is

$$\begin{pmatrix} \cos \theta & -\sin \theta & 0 \\ \sin \theta & \cos \theta & 0 \\ x \sin \theta \cos \theta - y \sin^2 \theta & -y \sin \theta \cos \theta - x \sin^2 \theta & 1 \end{pmatrix}$$

and may be rewritten as

$$\begin{pmatrix} \cos \theta & -\sin \theta & 0 \\ \sin \theta & \cos \theta & 0 \\ x' \sin \theta & x' \cos \theta - x & 1 \end{pmatrix}.$$

The differential $d\varphi$ sends the orthonormal basis (16) to

$$(\cos \theta, \sin \theta, x' \sin \theta), \quad (-\sin \theta, \cos \theta, x' \cos \theta), \quad (0, 0, 1).$$

These vectors at (x', y', z') are also orthonormal. $\qquad \square$

We deduce the following.

Proposition 12.5.4. *Every isometry of* Nil *preserves the line bundle
and induces an isometry of* $\mathbb{R}^2$. *We have*

$$0 \longrightarrow \mathbb{R} \longrightarrow \mathsf{Isom}^+(\mathsf{Nil}) \xrightarrow{p} \mathsf{Isom}(\mathbb{R}^2) \longrightarrow 0$$

where $\mathbb{R}$ *is the centre of* Nil. *The group* $\mathsf{Isom}^+(\mathsf{Nil})$ *has two components.*

Proof. The rotational isometries φ with angle θ introduced above
form a S^1-subgroup of $\mathsf{Isom}^+(\mathsf{Nil})$. The subgroups S^1 and Nil belong to the
component $\mathsf{Isom}_o^+(\mathsf{Nil}) < \mathsf{Isom}^+(\mathsf{Nil})$ containing the identity, which hence
has dimension at least $1 + 3 = 4$ and acts transitively on Nil. Stabilisers
have dimension at least $4 - 3 = 1$ and cannot have bigger dimension by
Proposition 6.2.15, otherwise Nil would have constant sectional curvature:
therefore $\dim(\mathsf{Isom}_o^+(\mathsf{Nil})) = 4$ and $\mathsf{Isom}_o^+(\mathsf{Nil})$ is generated by S^1 and Nil,
and it preserves the line bundle since S^1 and Nil do.

The subgroup $\mathsf{Isom}_o^+(\mathsf{Nil})$ preserves the orientation of the fibres and
of $\mathbb{R}^2$. We leave as an exercise the existence of another component of
$\mathsf{Isom}^+(\mathsf{Nil})$ which inverts the orientation of the fibres and of $\mathbb{R}^2$. $\qquad \square$
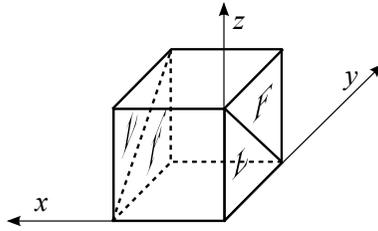
Figure 12.5. The Nil manifold $M = \text{Nil}/_\Gamma$ is obtained by identifying the faces of this unit cube as follows: the triangular faces are glued via the affine maps as shown by the labels, and the pairs of unlabelled opposite square faces are identified by translations. The cube is a fundamental domain for Γ.

Example 12.5.5. A manifold $M = \text{Nil}/_\Gamma$ modelled on Nil is described in Figure 12.5. The figure shows a fundamental domain for the group $\Gamma < \text{Isom}^+(\text{Nil})$ generated by the isometries

$$(x, y, z) \longmapsto (x + 1, y, z + y),$$

$$(x, y, z) \longmapsto (x, y + 1, z), \quad (x, y, z) \longmapsto (x, y, z + 1).$$

obtained by left-multiplication with the canonical basis of $\mathbb{R}^3$. Both the unit cube and M have volume one. The manifold M is clearly a torus bundle with monodromy $\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$ and is hence diffeomorphic to the Seifert manifold $\big(T, (1, 1)\big)$ by Exercise 11.4.11.

A group $\Gamma < \text{Isom}(X)$ is *cocompact* if $X/_\Gamma$ is compact. We prove that cocompact groups in $\text{Isom}(\mathbb{R}^2)$ do not lift.

Proposition 12.5.6. *Let* $\Gamma < \text{Isom}(\mathbb{R}^2)$ *be discrete and cocompact. There is no homomorphism* $f : \Gamma \to \text{Isom}^+(\text{Nil})$ *such that* $p \circ f = \text{id}$.

Proof. Up to taking a finite-index subgroup we may suppose that Γ is generated by two translations along independent vectors (x, y) and (x', y'). We prove that two lifts $\varphi, \varphi' \in \text{Isom}^+(\text{Nil})$ of these translations never commute, thus forbidding the existence of a homomorphism f.

If φ, $\varphi' \in \text{Nil}$, Exercise 12.5.1 gives $[\varphi, \varphi'] = (0, 0, xy' - x'y) \neq 0$. All the other lifts are of type $g\varphi, g'\varphi'$ for some $g, g' \in \mathbb{R}$, hence we get the same commutator. $\qquad\square$

Let $\text{Isom}_0^+(\text{Nil}) < \text{Isom}^+(\text{Nil})$ be the component containing e.

Exercise 12.5.7. We have

$$\big[\text{Isom}_0^+(\text{Nil}), \text{Isom}_0^+(\text{Nil})\big] = \mathbb{R}.$$

Therefore $\text{Isom}_0^+(\text{Nil})$ is nilpotent.

Corollary 12.5.8. *If a three-manifold M has a* Nil *geometry, then $\pi_1(M)$ is virtually nilpotent.*

Proof. It has a nilpotent subgroup of index at most two, because the subgroup $\mathrm{Isom}_0^+(\mathrm{Nil}) < \mathrm{Isom}^+(\mathrm{Nil})$ is nilpotent and has index two. □

12.5.4. Nil geometry. We classify the manifolds modelled on Nil.

Proposition 12.5.9. *A discrete group $\Gamma < \mathrm{Isom}^+(\mathrm{Nil})$ is cofinite if and only if both $p(\Gamma)$ and $\Gamma \cap \ker p$ are discrete and cofinite.*

Proof. We follow the proof of Proposition 12.4.4. Up to taking finite-index subgroups we may suppose that $\Gamma < \mathrm{Isom}_0^+(\mathrm{Nil})$. If $p(\Gamma)$ is discrete the proof of Proposition 12.4.4 applies also here; note that the formula (15) holds because the volume form in Nil is the standard one on $\mathbb{R}^3$.

If $p(\Gamma)$ is not discrete then that proof shows that there is a neighbourhood $U \subset \mathrm{Isom}^+(\mathrm{Nil})$ of e such that every two elements $f, g \in \Gamma$ projecting in $p(U)$ commute. (It is still true that $[f, g]$ depends only on $p(f)$ and $p(g)$ since $\ker p = \mathbb{R}$ is central.) We deduce again that the isometries in $p(\Gamma)$ commute: hence they are either rotations fixing the same point or translations.

In the former case Γ is not cofinite. In the latter, we get $\Gamma < \mathrm{Nil}$ acting as left-multiplication. Every two $f, g \in \Gamma$ projecting to $p(U)$ commute and hence project to parallel translations by Exercise 12.5.1. As in the proof of Proposition 12.4.4 we deduce that $p(\Gamma)$ preserves a line, and hence it is not cofinite. □

Corollary 12.5.10. *If the interior of a compact orientable manifold M admits a finite-volume complete* Nil *geometry then M is a closed Seifert manifold with $\chi = 0$ and $e \neq 0$.*

Proof. We have $\mathrm{int}(M) = \mathrm{Nil}/_\Gamma$ with Γ cofinite and Proposition 12.5.9 provides a Seifert fibration $M \to S$ over a finite-area orbifold $S = \mathbb{R}^2/_{p(\Gamma)}$. Finite-area flat orbifolds have $\chi(S) = 0$ and are closed (there are no cusps in flat geometry), hence M is closed.

We have $e \neq 0$, otherwise up to finite-index we would get $M = T \times S^1$ contradicting Proposition 12.5.6. □

We prove the converse.

Proposition 12.5.11. *If M is a closed Seifert manifold with $\chi = 0$ and $e \neq 0$, then M admits a* Nil *geometry.*

Proof. We have

$$M = \big(S, (p_1, q_1), \ldots, (p_k, q_k)\big).$$

Suppose S is a closed orientable surface of genus $g \geqslant 0$. In the following $\pi_1(S)$ is the orbifold fundamental group. We have:

$$\pi_1(S) = \langle a_1, b_1, \dots, a_g, b_g, c_1, \dots, c_k \mid [a_1, b_1] \cdots [a_g, b_g] c_1 \cdots c_k, c_i^{p_i} \rangle,$$

$$\pi_1(M) = \langle a_1, b_1, \dots, a_g, b_g, c_1, \dots, c_k, l \mid [a_1, b_1] \cdots [a_g, b_g] c_1 \cdots c_k, c_i^{p_i} l^{q_i},$$
$$[a_i, l], [b_i, l] \rangle.$$

We fix any flat structure on the orbifold S and get an injection $\pi_1(S) \to$ $\mathsf{Isom}^+(\mathbb{R}^2)$. We now want to lift this map to an injection $\pi_1(M) \to$ $\mathsf{Isom}^+(\mathsf{Nil})$, so that the resulting diagram commutes:

$$
\begin{array}{ccc}
\pi_1(M) & \dashrightarrow & \mathsf{Isom}^+(\mathsf{Nil}) \\
\downarrow & & \downarrow \\
\pi_1(S) & \longrightarrow & \mathsf{Isom}^+(\mathbb{R}^2)
\end{array}
$$

Recall that $\mathsf{Isom}_0^+(\mathsf{Nil})$ contains the group $\mathbb{R}$ of vertical translations: we use the multiplicative notation and indicate it as $\mathbb{R}_{>0}$. We note that $\mathbb{R}_{>0}$ is central in $\mathsf{Isom}_0^+(\mathsf{Nil})$ because it commutes with Nil and the isometries φ from Proposition 12.5.3, which altogether generate $\mathsf{Isom}_0^+(\mathsf{Nil})$.

We identify $\pi_1(S)$ with its image in $\mathsf{Isom}^+(\mathbb{R}^2)$. We lift arbitrarily a_i, b_i, c_i inside $\mathsf{Isom}^+(\mathsf{Nil})$ and pick $l \in \mathbb{R}_{>0}$. We get

$$[a_1, b_1] \cdots [a_g, b_g] c_1 \cdots c_k = e^\mu,$$
$$c_i^{p_i} l^{q_i} = e^{\lambda_i}$$

for some $\mu, \lambda_i \in \mathbb{R}$. To get a homomorphism $\pi_1(M) \to \mathsf{Isom}^+(\mathsf{Nil})$ we need $\mu = \lambda_i = 0$. To obtain that we change the lifts as $c_i' = e^{t_i} c_i$, $l' = e^u l$ to get

$$[a_1, b_1] \cdots [a_g, b_g] c_1' \cdots c_k' = e^{t_1 + \dots + t_k + \mu},$$
$$(c_i')^{p_i} (l')^{q_i} = e^{t_i p_i + u q_i + \lambda_i}.$$

We want

$$t_1 + \dots + t_k = -\mu,$$
$$t_i p_i + u q_i = -\lambda_i.$$

The determinant of the $(k + 1) \times (k + 1)$ coefficient matrix was already calculated in the proof of Proposition 10.3.15, and is

$$\pm \sum \frac{q_i}{p_i} (p_1 \cdots p_k) = \pm e \cdot p_1 \cdots p_k \neq 0.$$

Therefore the linear system has a unique solution $(t_1, \dots, t_k, u)$. The lift l' is necessarily non-trivial, otherwise we would get a lift $\pi_1(S) \to \mathsf{Isom}^+(\mathsf{Nil})$ that is excluded by Proposition 12.5.6. Therefore the lift is injective.

If S is non-orientable the proof is similar and left as an exercise. $\qquad \square$

12.6. $\widetilde{\mathrm{SL}}_2$ geometry

The $\widetilde{\mathrm{SL}}_2$ geometry is similar to Nil, but it is now a $\mathbb{R}$-bundle over $\mathbb{H}^2$. The geometry is again fully carried by the Lie group $\widetilde{\mathrm{SL}}_2$, which is the universal cover of both $\mathrm{SL}_2(\mathbb{R})$ and $\mathbb{PSL}_2(\mathbb{R})$. It is convenient to identify $\mathbb{PSL}_2(\mathbb{R})$ with the unit tangent bundle of $\mathbb{H}^2$.

12.6.1. The unit tangent bundle. The tangent bundle TM of a Riemannian n-manifold M has a natural Riemannian structure, which we briefly introduce.

The tangent space at any point $x \in TM$ splits into a *vertical* subspace $V_x = \ker dp_x$ where $p \colon TM \to M$ is the projection, and a *horizontal* subspace H_x determined by the metric tensor on M as follows: use the parallel transport to move the tangent vector x along all geodesics exiting from $p(x)$; the result is a small n-surface in TM containing x and we set H_x to be its tangent space at x.

To define a metric on TM we impose that V_x and H_x be orthogonal and we give to each space the metric of T_xM, via the natural identification $V_x = T_xM$ and via the isomorphism $dp_x \colon H_x \to T_xM$.

The *unit tangent bundle* $UM \subset TM$ consists of all unitary tangent vectors and inherits a Riemannian structure. Every isometry $f \colon M \to M$ induces an isometry $df \colon UM \to UM$.

12.6.2. The space $U\mathbb{H}^2$. We now focus on the case $M = \mathbb{H}^2$ we are interested in. Parallel transport was defined explicitly in Section 2.4.4. We represent $\mathbb{H}^2$ using the upper half-plane model $H^2 = \{\Im z > 0\}$, so that $TH^2 = H^2 \times \mathbb{C}$ and $UH^2 = H^2 \times S^1$ has coordinates (z, θ). The tangent space $T_{(z,\theta)}UH^2$ is naturally identified with $\mathbb{C} \times \mathbb{R} = \mathbb{R}^3$.

Lemma 12.6.1. *The metric tensor of UH^2 at (z, θ) is*
$$\begin{pmatrix} 2y^{-2} & 0 & y^{-1} \\ 0 & y^{-2} & 0 \\ y^{-1} & 0 & 1 \end{pmatrix}$$
where $z = x + iy$.

Proof. We first consider the disc model D^2 of $\mathbb{H}^2$ with $TD^2 = D^2 \times \mathbb{C}$ and $T_{(z,v)}TD^2 = \mathbb{C} \times \mathbb{C}$. We focus at a point $(0, v)$ and determine the decomposition
$$T_{(0,v)} = H_{(0,v)} \oplus V_{(0,v)}.$$
We have $V_{(0,v)} = 0 \times \mathbb{C}$ and we now determine $H_{(0,v)}$. Every geodesic through 0 is a Euclidean line l and the parallel transport of $v \in T_0D^2$ along l forms a constant angle with l: the parallel transport of v at $z \in l$ is just the rescaled vector $(1 - |z|^2)v$. Parallel transports of v along lines passing through 0 form a surface, which is the graph of the function $f(z) = (1 - |z|^2)v$, whose tangent plane at 0 is $H_{(0,v)} = \mathbb{C} \times 0$ since $\frac{\partial f}{\partial x} = \frac{\partial f}{\partial y} = 0$ at 0.

We have discovered that the horizontal and vertical planes at $(0, v)$ are just the coordinate ones of $T_{(0,v)} = \mathbb{C} \times \mathbb{C}$. By definition these are orthogonal and inherit a metric tensor from that of $T_0 D^2$, which is 4 times the Euclidean one. Therefore the metric tensor at $T_{(0,v)} = \mathbb{C}^2$ is 4 times the Euclidean one.

We now turn to the unitary sub-bundle $UD^2 = D^2 \times S^1$. We use the natural identifications $T_{(0,\theta)} UD^2 = \mathbb{C} \times \mathbb{R} = \mathbb{R}^3$ and we deduce that the vectors $(1, 0, 0)$, $(0, 1, 0)$, $(0, 0, 1) \in T_{(0,\theta)}$ are orthogonal and have norm $2, 2, 1$ respectively.

The Möbius transformation

$$f(z) = \frac{z - i}{iz - 1}$$

is an isometry between the two models D^2 and H^2 that sends 0 to i. It is holomorphic and its complex derivative is

$$f'(z) = \frac{-2}{(iz - 1)^2}.$$

The isometry $f: D^2 \to H^2$ induces an isometry $f_*: UD^2 \to UH^2$, that is $f_*: D^2 \times S^1 \to H^2 \times S^1$, which is as follows:

$$f_*(z, \theta) = \left(\frac{z - i}{iz - 1}, \theta + \arg(f'(z)) \right)$$

$$= \left(\frac{z - i}{iz - 1}, \theta + \pi - 2\arg(iz - 1) \right)$$

$$= \left(\frac{z - i}{iz - 1}, \theta + \pi - 2\Im \log(iz - 1) \right).$$

In particular $f_*(0, \theta) = (i, \theta + \pi)$. Recall that the Jacobians of a holomorphic g and of $\log g$ are

$$Jg = \begin{pmatrix} \Re g' & -\Im g' \\ \Im g' & \Re g' \end{pmatrix}, \quad J(\log g) = \begin{pmatrix} \Re \frac{g'}{g} & -\Im \frac{g'}{g} \\ \Im \frac{g'}{g} & \Re \frac{g'}{g} \end{pmatrix}.$$

The differential of f_* at the point (z, θ) is hence

$$(df_*)_{(z,\theta)} = \begin{pmatrix} \Re \frac{-2}{(iz-1)^2} & -\Im \frac{-2}{(iz-1)^2} & 0 \\ \Im \frac{-2}{(iz-1)^2} & \Re \frac{-2}{(iz-1)^2} & 0 \\ -2\Im \frac{1}{z+i} & -2\Re \frac{1}{z+i} & 1 \end{pmatrix}.$$

In particular at $z = 0$ we get

$$(df_*)_{(0,\theta)} = \begin{pmatrix} -2 & 0 & 0 \\ 0 & -2 & 0 \\ 2 & 0 & 1 \end{pmatrix}.$$

Since this is an isometry, the three image vectors $(-2, 0, 2)$, $(0, -2, 0)$, and $(0, 0, 1)$ are orthogonal with norm $2, 2, 1$, hence $(1, 0, -1)$, $(0, 1, 0)$, and

$(0, 0, 1)$ form an orthonormal basis at $T_{(i, \theta + \pi)} UH^2$ and the metric tensor there is

$$\begin{pmatrix} 2 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}.$$

The tensor is independent of θ. To find its value at (z, θ) for a generic $z \in H^2$ it suffices to transport it via an isometry that sends i to z. Every such isometry is a composition of a horizontal translation (which does not change the tensor) and a dilation $g(z) = \lambda z$ with $\lambda > 0$. We have $g'(z) = \lambda$ and $g_*(z, \theta) = (\lambda z, \theta)$, therefore dg_* sends the above orthonormal basis at $T_{(i, \theta)}$ to the basis $(\lambda, 0, -1)$, $(0, \lambda, 0)$, $(0, 0, 1)$. The metric tensor at $(\lambda i, \theta)$ is hence

$$\begin{pmatrix} 2\lambda^{-2} & 0 & \lambda^{-1} \\ 0 & \lambda^{-2} & 0 \\ \lambda^{-1} & 0 & 1 \end{pmatrix}.$$

The proof is complete.   □

At a point $(x, y, \theta) \in UH^2$, an orthonormal basis is

(17) $\qquad\qquad (y, 0, -1), \quad (0, y, 0), \quad (0, 0, 1).$

The planes orthogonal to $(0, 0, 1)$ form a contact structure similar to the one drawn in Figure 12.4.

We remark that the metric tensor at (x, y, θ) has determinant y^{-4}, like in the product metric $H^2 \times S^1$: therefore the volume form in UH^2 is the same as in the product metric (although the metric tensor is not).

The non-zero Christoffel symbols at (x, y, θ) are:

$$\Gamma_{12}^1 = \Gamma_{21}^1 = -\frac{3}{2y}, \quad \Gamma_{23}^1 = \Gamma_{32}^1 = -\frac{1}{2},$$

$$\Gamma_{11}^2 = \frac{2}{y}, \quad \Gamma_{13}^2 = \Gamma_{31}^2 = \frac{1}{2}, \quad \Gamma_{22}^2 = -\frac{1}{y},$$

$$\Gamma_{12}^3 = \Gamma_{21}^3 = \frac{1}{y^2}, \quad \Gamma_{23}^3 = \Gamma_{32}^3 = \frac{1}{2y}.$$

The Ricci tensor is

$$R_{ij} = \frac{1}{2} \cdot \begin{pmatrix} -2y^{-2} & 0 & y^{-1} \\ 0 & -3y^{-2} & 0 \\ y^{-1} & 0 & 1 \end{pmatrix}.$$

When $y = 1$ we can represent the Ricci tensor in the orthonormal basis (17) and get a diagonal matrix with values $-\frac{3}{2}$, $-\frac{3}{2}$, $\frac{1}{2}$. For a vector $v \in T_{(z, \theta)} U\mathbb{H}^2$, recall that $R_{ij} v_i v_j$ is twice the average value of the sectional curvatures of the planes containing v: this average value ranges from $-\frac{3}{4}$ to $\frac{1}{4}$ and is maximal when $v = (0, 0, \pm 1)$. This holds when $y = 1$ and hence for any $z \in U\mathbb{H}^2$ by acting via isometries of $\mathbb{H}^2$.

12.6.3. $\widetilde{SL_2}$ geometry. Let $\widetilde{SL_2}$ be the universal cover of $SL_2 = SL_2(\mathbb{R})$. As a universal cover of a Lie group, it is also a Lie group. We have coverings

$$\widetilde{SL_2} \longrightarrow SL_2 \longrightarrow \mathbb{PSL}_2(\mathbb{R}) = \mathsf{Isom}^+(\mathbb{H}^2).$$

The group $\mathbb{PSL}_2(\mathbb{R})$ acts freely and transitively on $U\mathbb{H}^2$ and hence we can identify $\mathbb{PSL}_2(\mathbb{R})$ with $U\mathbb{H}^2$. With this identification $\mathbb{PSL}_2(\mathbb{R})$ inherits a left-invariant Riemannian metric that lifts to a left-invariant Riemannian metric on the Lie group $\widetilde{SL_2}$.

In the previous section we have identified $U\mathbb{H}^2$ with $H^2 \times S^1$ and we can likewise identify $\widetilde{SL_2}$ with $H^2 \times \mathbb{R}$, and get an explicit metric tensor from Lemma 12.6.1.

Since $U\mathbb{H}^2$ is a circle bundle over $\mathbb{H}^2$, likewise $\widetilde{SL_2}$ is a line bundle over $\mathbb{H}^2$. The group $\mathbb{R}$ acts isometrically on $\widetilde{SL_2}$ by translating every fibre.

Proposition 12.6.2. *Every isometry of $\widetilde{SL_2}$ preserves the line bundle and induces an isometry of $\mathbb{H}^2$. We have*

$$0 \longrightarrow \mathbb{R} \longrightarrow \mathsf{Isom}^+(\widetilde{SL_2}) \overset{p}{\longrightarrow} \mathsf{Isom}(\mathbb{H}^2) \longrightarrow 0.$$

The group $\mathsf{Isom}^+(\widetilde{SL_2})$ has two components.

Proof. The groups $\mathbb{R}$ and $\widetilde{SL_2}$ belong to the component $\mathsf{Isom}_o^+(\widetilde{SL_2})$ containing e which has dimension at least $1 + 3 = 4$. We conclude as in the proof of Proposition 12.5.4. □

Exercise 12.6.3. The unit tangent bundle of a finite-area complete hyperbolic surface is naturally a manifold modelled on $\widetilde{SL_2}$.

As for Nil, cocompact groups do no lift.

Proposition 12.6.4. *Let $\Gamma < \mathsf{Isom}(\mathbb{H}^2)$ be discrete and cocompact. There is no homomorphism $f \colon \Gamma \to \mathsf{Isom}^+(\widetilde{SL_2})$ such that $p \circ f = \mathrm{id}$.*

Proof. Up to taking a finite index subgroup we may suppose that Γ acts freely and hence $S = \mathbb{H}^2/_\Gamma$ is a closed hyperbolic surface. If Γ lifts, consider the group $G < \mathsf{Isom}^+(\widetilde{SL_2})$ generated by $f(\Gamma)$ and $2\pi \in \mathbb{R}$, isomorphic to $\Gamma \times \mathbb{Z}$. The quotient $\widetilde{SL_2}/_G$ is naturally the unit tangent bundle of S and is hence the Seifert manifold $(S, (1, \chi(S)))$ by Remark 10.2.8. Its fundamental group is however not a product $\Gamma \times \mathbb{Z}$ by Exercise 10.2.9. □

Proposition 12.6.5. *A discrete group $\Gamma < \mathsf{Isom}^+(\widetilde{SL_2})$ is cofinite if and only if both $p(\Gamma)$ and $\Gamma \cap \ker p$ are discrete and cofinite.*

Proof. Same proof as Proposition 12.4.4, with a minor variation: if $p(\Gamma)$ is not discrete, up to taking an index-two subgroup we suppose that $p(\Gamma) < \mathsf{Isom}^+(\mathbb{H}^2)$; the $\mathbb{R}$-action commutes with Γ and $[f, g]$ depends only on $p(f)$ and $p(g)$, so that proof applies. □

We now classify the manifolds having a $\widetilde{\mathrm{SL}}_2$ geometry.

Corollary 12.6.6. *If the interior of a compact orientable manifold M admits a finite-volume complete $\widetilde{\mathrm{SL}}_2$ geometry then M is a Seifert manifold with $\chi < 0$. If M is closed then $e \neq 0$.*

Proof. We have $\mathrm{int}(M) = \widetilde{\mathrm{SL}}_2/_\Gamma$ with Γ cofinite. Proposition 12.6.5 furnishes a Seifert fibration $M \to S$ over the finite-area orbifold $S = \mathbb{H}^2/_{p(\Gamma)}$. If M is closed we get $e \neq 0$: if not, up to taking a finite index subgroup we would get $M = S \times S^1$ contradicting Proposition 12.6.4. $\quad\square$

We now prove the converse.

Proposition 12.6.7. *If M is a Seifert manifold with $\chi < 0$ and either $\partial M \neq \varnothing$ or $e \neq 0$, the interior of M admits a finite-volume $\widetilde{\mathrm{SL}}_2$ geometry.*

Proof. If M is closed we apply the proof of Proposition 12.5.11. If $\partial M \neq \varnothing$ the presentations of $\pi_1(S)$ and $\pi_1(M)$ are as described there, except that they do not contain the relator $c_i^{p_i}$ whenever c_i represents a boundary component of M. So we have less constraints and we easily see that a solution to the final linear problem exists also in this case. $\quad\square$

12.7. Sol geometry

The Sol geometry is the least symmetric one among the eight. It has a bundle structure, but with a one-dimensional basis: it is a $\mathbb{R}^2$-bundle over $\mathbb{R}$. Again, the geometry is fully governed by a Lie group Sol.

12.7.1. The Lie group Sol. The Lie group Sol is the space $\mathbb{R}^3$ equipped with the following operation

$$(x, y, z) \cdot (x', y', z') = (x + e^{-z}x', y + e^z y', z + z').$$

Exercise 12.7.1. We have

$$(x, y, z)^{-1} = (-xe^z, -ye^{-z}, -z),$$

$$[(x, y, z), (x', y', z')] = \big(x(1-e^{-z'})-x'(1-e^{-z}), y(1-e^{z'})-y'(1-e^z), 0\big),$$

$$[(x, y, z), (x', y', 0)] = \big(-x'(1-e^{-z}), -y'(1-e^z), 0\big).$$

The subgroup $\mathbb{R}^2$ consisting of all elements $(x, y, 0)$ is normal in Sol and by setting $p(x, y, z) = z$ we get an exact sequence

$$0 \longrightarrow \mathbb{R}^2 \longrightarrow \mathrm{Sol} \overset{p}{\longrightarrow} \mathbb{R} \longrightarrow 0.$$

Therefore Sol is a plane bundle over $\mathbb{R}$. Exercise 12.7.1 implies the following.

Exercise 12.7.2. We have $[\mathrm{Sol}, \mathrm{Sol}] = \mathbb{R}^2$ and hence Sol is solvable. However $[\mathrm{Sol}, \mathbb{R}^2] = \mathbb{R}^2$ and hence Sol is not nilpotent.

We define a Riemannian metric on Sol by assigning the scalar product

$$\begin{pmatrix} e^{2z} & 0 & 0 \\ 0 & e^{-2z} & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

to the point (x, y, z). The metric is left-invariant and every plane $z = k$ is isometric to the Euclidean $\mathbb{R}^2$. This is the geometry with the smallest amount of symmetries.

We remark that the metric tensor has unit determinant at every point: therefore the volume form on Sol is the standard one on $\mathbb{R}^3$.

The non-zero Christoffel symbols at (x, y, z) are

$$\Gamma_{13}^1 = \Gamma_{31}^1 = 1,$$

$$\Gamma_{23}^2 = \Gamma_{32}^2 = -1,$$

$$\Gamma_{11}^3 = -e^{2z}, \quad \Gamma_{22}^3 = e^{-2z}.$$

The Ricci tensor is

$$\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -2 \end{pmatrix}.$$

The average value of the sectional curvatures of the planes containing $v \in T_{(x,y,z)}$Sol ranges from -1 to 0, and is minimal when $v = (0, 0, \pm 1)$.

12.7.2. Sol geometry. We start with a simple exercise.

Exercise 12.7.3. The eight maps

$$(x, y, z) \mapsto (\pm x, \pm y, z), \qquad (x, y, z) \mapsto (\pm y, \pm x, -z)$$

are isometries and form the dihedral group D_8. The orientation-preserving ones form the subgroup $D_4 = \mathbb{Z}_2 \times \mathbb{Z}_2$.

Let $\text{Isom}_*(\mathbb{R}^2) < \text{Isom}(\mathbb{R}^2)$ be the subgroup consisting of all maps $v \mapsto \pm v + b$. It has two components, one being the translations $\mathbb{R}^2$.

Proposition 12.7.4. *Every isometry of* Sol *preserves the plane bundle and induces an isometry of* $\mathbb{R}$. *We have*

$$0 \longrightarrow \text{Isom}_*(\mathbb{R}^2) \longrightarrow \text{Isom}^+(\text{Sol}) \overset{p}{\longrightarrow} \text{Isom}(\mathbb{R}) \longrightarrow 0.$$

The group $\text{Isom}^+(\text{Sol})$ *has four components, one of which is* Sol *acting by left-multiplication.*

Proof. The group Sol acts transitively and freely on Sol itself, and to conclude it suffices to check that the stabiliser of the origin 0 is the dihedral D_4 described in Exercise 12.7.3.

The Ricci tensor tells us that an isometry fixing a point also fixes the vertical axis and the horizontal plane. Therefore the vertical unitary constant vector field $X = (0, 0, 1)$ is preserved up to sign by any isometry of Sol. The covariant differentiation $v \mapsto \nabla_v X$ defines an endomorphism of

T_pSol for all $p \in$ Sol. We have $\nabla_{e_i} X = \Gamma^k_{i3} e_k$ and therefore the endomorphism is

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

The three coordinate axis are precisely the eigenvectors of the endomorphism, and being intrinsically defined they are preserved by every isometry. Therefore the orientation-preserving stabiliser of a point is D_4. □

Proposition 12.7.5. *A discrete group* $\Gamma < \mathrm{Isom}^+(\mathrm{Sol})$ *is cofinite if and only if both* $p(\Gamma)$ *and* $\Gamma \cap \ker p$ *are discrete and cofinite.*

Proof. If $p(\Gamma)$ is discrete we get

$$\mathrm{Vol}\big(\mathrm{Sol}/_\Gamma\big) = \mathrm{Length}(\mathbb{R}/_{p(\Gamma)}) \times \mathrm{Area}\big(\mathbb{R}^2/_{\Gamma \cap \ker p}\big).$$

This formula is proved as above by taking a fundamental domain for $p(\Gamma)$. We now prove that $p(\Gamma)$ is in fact discrete (since Γ is).

Up to taking an index-four subgroup we suppose that $\Gamma < \mathrm{Sol}$. If $\gamma \in \mathrm{Sol}$ does not lie in $\mathbb{R}^2$, then it fixes a vertical line. To prove that, note that $\gamma = (x, y, z)$ acts on $\mathbb{R}^3$ as an affine transformation, which permutes the vertical lines and acts on $\mathbb{R}^2$ as $(x', y') \mapsto (x + e^{-z}x', y + e^z y')$. If $z \neq 0$ this map has a fixed point (because 1 is not an eigenvalue of its linearisation).

If Γ is abelian, either it is contained in $\mathbb{R}^2$ or it fixes a vertical line: in both cases we get a discrete $p(\Gamma)$. If Γ is non-abelian, then $[\Gamma, \Gamma] < \mathbb{R}^2$ is non-trivial so Γ contains a non-trivial element $\gamma \in \mathbb{R}^2$ and another $\eta \notin \mathbb{R}^2$. The elements γ and $\eta\gamma\eta^{-1}$ are both in $\mathbb{R}^2$ and independent, hence $\mathbb{R}^2/_{\Gamma \cap \ker p}$ is compact. This implies easily that $p(\Gamma)$ is discrete (since Γ is). □

We now classify the manifolds modelled on Sol. Recall that every semi-bundle is doubly covered by a canonical bundle. A torus bundle is *of Anosov type* if its monodromy is Anosov. A torus semi-bundle is of Anosov type if its double-covering is.

Proposition 12.7.6. *The interior of a compact orientable manifold M admits a finite-volume complete* Sol *geometry if and only if it is a torus (semi-)bundle of Anosov type.*

Proof. If $\mathrm{int}(M) = \mathrm{Sol}/_\Gamma$ has finite volume, then both $\Gamma \cap \ker p$ and $p(\Gamma)$ are discrete and cofinite. The horizontal foliation of Sol into Euclidean planes $z = k$ is preserved by Γ and projected to a surface bundle over the 1-orbifold $\mathbb{R}/_{p(\Gamma)}$, which is either S^1 or an interval. The fibres are flat and have finite area, hence they are tori or Klein bottles.

If $p(\Gamma)$ acts by translations, then M is a torus bundle with an Anosov monodromy $\pm\big(\begin{smallmatrix} e^{-z} & 0 \\ 0 & e^z \end{smallmatrix}\big)$ for some $z > 0$. If M is a semi-bundle, this holds on its double-covering.

Conversely, let M_A be a torus bundle with Anosov monodromy $A \in$ $SL_2(\mathbb{Z})$. We have $M_A = \mathbb{R}^3/_\Gamma$ where Γ is generated by the affine maps

$$(x, y, z) \mapsto (x+1, y, z), \quad (x, y, z) \mapsto (x, y+1, z), \quad (\bar{x}, z) \mapsto (A\bar{x}, z+h)$$

where $\bar{x} = (x, y)$ and $h > 0$ is any positive number. A linear isomorphism ψ of $\mathbb{R}^2$ conjugates A into a diagonal matrix $\pm\left(\begin{smallmatrix} e^{-a} & 0 \\ 0 & e^a \end{smallmatrix}\right)$ for some $a > 0$ and transforms the generators into

$$(\bar{x}, z) \mapsto (\bar{x}+\bar{v}, z), \ (\bar{x}, z) \mapsto (\bar{x}+\bar{w}, z), \ (x, y, z) \mapsto (\pm e^{-a}x, \pm e^a y, z+h)$$

where $\bar{v} = \psi(1, 0)$ and $\bar{w} = \psi(0, 1)$. If $\mathrm{tr}A > 0$ and $h = a$ these maps are left-multiplications by the following elements of Sol:

$$(\bar{v}, 0), \quad (\bar{w}, 0), \quad (\bar{0}, a)$$

where $\bar{0} = (0, 0)$, and we are done. If $\mathrm{tr}A < 0$ we compose the third generator with the isometry $(x, y, z) \mapsto (-x, -y, z)$.

If M is a semi-bundle double-covered by M_A, we represent it as $\mathbb{R}^3/_{\Gamma'}$ where Γ' is generated by the group Γ representing M_A plus the element

$$(x, y, z) \mapsto \left(x + \frac{1}{2}, -y, -z\right)$$

which will be transformed via ψ into an isometry of Sol.   □

Corollary 12.7.7. *If a closed three-manifold M has a* Sol *geometry, then $\pi_1(M)$ is virtually solvable but not virtually nilpotent.*

Proof. It is virtually solvable because Sol has finite index in $\mathrm{Isom}^+(\mathrm{Sol})$. The proof of Proposition 12.7.6 shows that up to finite index $M = \mathrm{Sol}/_\Gamma$ and Γ is generated by

$$(\bar{v}, 0), \quad (\bar{w}, 0), \quad (\bar{0}, a).$$

Exercise 12.7.1 implies that $[(\bar{0}, a), (\bar{v}, 0)] = (\bar{v}', 0)$ for some $\bar{v}' \neq 0$ and hence Γ is not nilpotent. This holds for any Γ of this kind, and therefore Γ is not virtually nilpotent, either.   □

12.8. Summary

We have proudly completed the proof of Theorem 12.1.1. We state that theorem again here:

Theorem 12.8.1. *A closed orentable 3-manifold has a geometric structure modelled on one of the following six geometries:*

$$S^3, \ \mathbb{R}^3, \ S^2 \times \mathbb{R}, \ \mathbb{H}^2 \times \mathbb{R}, \ \mathrm{Nil}, \ \widetilde{SL_2}$$

if and only if it is a Seifert manifold of the appropriate commensurability class, as prescribed by Table 12.1. It has a Sol *geometric structure if and only if it is a torus (semi-)bundle of Anosov type.*

We can deduce that the eight geometries are mutually exclusive:

Proposition 12.8.2. *Two closed 3-manifolds admitting different geometries are not diffeomorphic, and not even commensurable.*

Proof. We already know that the six Seifert geometries form six distinct commensurability classes. A manifold of type Sol is not Seifert by Proposition 11.4.14. A closed hyperbolic manifold is neither Seifert nor Sol because its fundamental group is infinite and does not contain an infinite normal abelian subgroup by Corollary 5.1.10 (as opposite to the other geometries, see Proposition 10.4.3).

Two manifolds having different geometries cannot be commensurable, since a common finite cover would inherit both geometries.   □

In fact one can tell the geometry of the manifold directly from its fundamental group.

Proposition 12.8.3. *Let M be a closed manifold modelled on one $\mathbb{X}$ of the eight geometries; this flowchart shows how to determine $\mathbb{X}$ from the fundamental group $\pi_1(M)$:*

- *if $\pi_1(M)$ is finite, then $\mathbb{X} = S^3$; otherwise*
- *if $\pi_1(M)$ is virtually cyclic, then $\mathbb{X} = S^2 \times \mathbb{R}$; otherwise*
- *if $\pi_1(M)$ is virtually abelian, then $\mathbb{X} = \mathbb{R}^3$; otherwise*
- *if $\pi_1(M)$ is virtually nilpotent, then $\mathbb{X} = $ Nil; otherwise*
- *if $\pi_1(M)$ is virtually solvable, then $\mathbb{X} = $ Sol; otherwise*
- *if $\pi_1(M)$ contains a normal cyclic group K, then:*
  - *if a finite-index subgroup of the quotient lifts, $\mathbb{X} = \mathbb{H}^2 \times \mathbb{R}$,*
  - *otherwise $\mathbb{X} = \widetilde{SL_2}$;*
- *otherwise $\mathbb{X} = \mathbb{H}^3$.*

Proof. The virtually abelian cases were settled in Section 10.3.10. The Nil and Sol geometries were considered by Corollaries 12.5.8 and 12.7.7.

If $\mathbb{X} = \mathbb{H}^2 \times \mathbb{R}$ or $\widetilde{SL_2}$ then $\pi_1(M)$ surjects onto $\pi_1(S)$ for some closed hyperbolic surface S. This implies that $\pi_1(M)$ is not solvable, for otherwise (by Propositions 1.4.3 and 1.4.10) the group $\pi_1(S)$ would also be, and hence it would contain a normal cyclic subgroup, contradicting Corollary 5.1.10.

If $\mathbb{X} = \widetilde{SL_2}$ then $\pi_1(S) = \pi_1(M)/_K$ and no finite-index subgroup of $\pi_1(S)$ can lift to $\pi_1(M)$ by Proposition 12.6.4.

Finally, if $\mathbb{X} = \mathbb{H}^3$ then $\pi_1(M)$ contains no normal cyclic group K by Corollary 5.1.10 again.   □

The boundary case is slightly different. Of the eight geometries, only three produce non-compact finite-volume complete orientable manifolds:

$$\mathbb{H}^3, \ \mathbb{H}^2 \times \mathbb{R}, \ \widetilde{SL_2}.$$

These non-compact finite-volume manifolds are diffeomorphic to the interior of a compact manifold M with boundary consisting of tori.

Theorem 12.8.4. *Let M be a compact orientable 3-manifold with boundary consisting of tori. The following are equivalent:*

- *M has a finite-volume complete $\mathbb{H}^2 \times \mathbb{R}$ structure,*
- *M has a finite-volume complete $\widetilde{\mathrm{SL}_2}$ structure,*
- *M is Seifert with $\chi < 0$.*

In contrast with the closed case, a non-closed M may admit two different geometric structures of Seifert type. On the other hand, a manifold M cannot admit both a hyperbolic and a Seifert structure.

12.9. The geometrisation conjecture

We can finally state the famous Geometrisation Conjecture, proposed by Thurston in 1982 and proved twenty years later by Perelman in 2002.

12.9.1. Statement and main consequences. We say that a compact 3-manifold with (possibly empty) boundary consisting of tori is *geometric* if its interior has a finite-volume complete geometric structure modelled on one of the eight geometries:

$$S^3, \ \mathbb{R}^3, \ \mathbb{H}^3, \ S^2 \times \mathbb{R}, \ \mathbb{H}^2 \times \mathbb{R}, \ \mathrm{Nil}, \ \mathrm{Sol}, \ \widetilde{\mathrm{SL}_2}.$$

The following conjecture was formulated by Thurston in 1982:

Conjecture 12.9.1 (Geometrisation Conjecture). *Let M be an irreducible orientable compact 3-manifold with (possibly empty) boundary consisting of tori. Every block of the geometric decomposition of M is geometric.*

The conjecture has been proved by Perelman in 2002 and its proof goes very very far from the scope of this book. It is however quite easy to deduce important consequences from it.

Conjecture 12.9.2 (Poincaré conjecture). *Every closed simply connected three-manifold M is diffeomorphic to S^3.*

Proof using geometrisation. Via the prime decomposition we may restrict to the case M is prime, hence irreducible. The group $\pi_1(M)$ is trivial and hence does not contain $\mathbb{Z} \times \mathbb{Z}$: every torus in M is thus compressible and the geometric decomposition is trivial. By geometrisation M is itself geometric. The only geometry with finite fundamental groups is S^3, and hence $M = S^3/_\Gamma$ is elliptic. Since M is simply connected, the group $\Gamma = \pi_1(M)$ is trivial and hence $M = S^3$. $\qquad\square$

Conjecture 12.9.3 (Elliptisation). *Every closed 3-manifold M with finite $\pi_1(M)$ is elliptic.*

Proof using geometrisation. Same proof as above. Note that this is *not* a consequence of Poincaré conjecture in general, for a manifold covered by S^n needs not to be elliptic a priori. $\qquad\square$

Conjecture 12.9.4 (Hyperbolisation). *Every closed irreducible 3-manifold M with infinite $\pi_1(M)$ not containing $\mathbb{Z} \times \mathbb{Z}$ is hyperbolic.*

Proof using geometrisation. Since $\pi_1(M)$ does not contain $\mathbb{Z} \times \mathbb{Z}$ every torus is compressible and the geometric decomposition of M is trivial. By geometrisation M is geometric. Its geometry is not S^3 since $\pi_1(M)$ is infinite, and is not $S^2 \times \mathbb{R}$ since M is irreducible. In the other Seifert geometries and in Sol the fundamental group $\pi_1(M)$ always contain a $\mathbb{Z} \times \mathbb{Z}$ (there is always a finite covering containing an incompressible torus). $\square$

Corollary 12.9.5. *Let $\tilde{M} \to M$ be a finite covering. If $\tilde{M}$ is geometric, then M also is (with the same geometry).*

Proof using geometrisation. Note that this is stronger than Conjecture 12.9.3. Since $\tilde{M}$ is geometric, it is irreducible and hence also M is. Proposition 11.5.7 implies that the geometric decomposition of M is trivial, and by geometrisation M is geometric. $\square$

Concerning bounded manifolds, we get the following simple statement.

Corollary 12.9.6. *Every simple compact manifold M bounded by a non-empty collection of tori is hyperbolic.*

Proof using geometrisation. Being simple, it is geometric. Seifert manifolds with boundary are never simple. $\square$

That statement is not true for closed manifolds, because many Seifert manifolds fibering over S^2 with at most 3 singular fibres are simple.

12.9.2. Surface bundles. There is a nice way to formulate geometrisation for surface bundles. We start with the much simpler torus case.

Proposition 12.9.7. *Let M_A be a torus bundle with monodromy $A \neq \pm I$. The following holds:*

- *if $|\mathrm{tr}A| < 2$, i.e. A has finite order, then M_A is flat;*
- *if $|\mathrm{tr}A| = 2$, i.e. A is reducible, then M_A is Nil;*
- *if $|\mathrm{tr}A| > 2$, i.e. A is Anosov, then M_A is Sol.*

Proof. Use Proposition 11.4.14. $\square$

We now turn to the generic case. Let Σ be a closed orientable surface with $\chi(\Sigma) < 0$.

Theorem 12.9.8. *Let M_ψ be a surface bundle with fibre Σ and monodromy $\psi \in \mathrm{MCG}(\Sigma)$. The following holds:*

- *if ψ has finite order, then M_ψ has a $\mathbb{H}^2 \times \mathbb{R}$ geometry,*
- *if ψ is reducible, then M_ψ contains an essential torus,*
- *if ψ is pseudo-Anosov, then M_ψ is hyperbolic.*

Proof using geometrisation. Use Proposition 11.4.15. $\square$

Theorem 12.9.8 was initially proved by Thurston in the 1980s. In the same years Thurston also proved the geometrisation conjecture for all Haken 3-manifolds. Before Perelman's proof the conjecture was open "only" in the non-Haken case, and it naturally split in three parts: the Conjectures 12.9.2, 12.9.3, and 12.9.4. Perelman's proof certifies geometrisation in all cases with a unified technique.

12.9.3. References. This chapter contains many technical proofs. Most of them were taken from Scott [**52**], that is the standard reference on the geometrisation of Seifert manifolds, and Thurson's book [**57**]. We have also consulted a nice survey of Bonahon [**8**], that contains in particular the suggestion of calculating the Ricci tensors in the non-product geometries. Thurson's geometrisation conjecture appears in [**58**] in 1982, while Perelman's proof consists of three papers that he sent to the arXiv in 2003 [**45, 46, 47**].

Mostow rigidity theorem

We have defined in Chapter 7 the Teichmüller space $\text{Teich}(S_g)$ of a genus-g closed orientable surface S_g as the space of all the hyperbolic metrics on S_g, considered up to isometries isotopic to the identity; we have then proved that $\text{Teich}(S_g)$ is homeomorphic to $\mathbb{R}^{6g-6}$ using the Fenchel–Nielsen coordinates.

This definition of $\text{Teich}(M)$ actually applies to any closed hyperbolic manifold M, and we show here a striking difference between the dimensions two and three: if $\dim M = 3$ then $\text{Teich}(M)$ is a single point. This strong result is known as the *Mostow rigidity Theorem*.

The impact of Mostow's rigidity on our knowledge of 3-dimensional topology cannot be overestimated. Thanks to this theorem every geometric information on a given closed hyperbolic three-manifold M like its volume, geodesic spectrum, etc. is promoted to a *topological* invariant of M, that is it depends on the differentiable structure of M only. In its strongest version, Mostow's rigidity says that the hyperbolic metric of M is fully determined by the group $\pi_1(M)$ alone.

We expose here Gromov's proof of Mostow's rigidity, which uses hyperbolic tetrahedra and introduces a nice invariant on closed manifolds of any dimension, called the *simplicial volume*.

13.1. Volume of tetrahedra

The volume of an ideal hyperbolic tetrahedron is a simple (but integral) formula that involves the *Lobachevsky function*: we now prove this formula here. As a consequence we show that the regular ideal tetrahedron is the hyperbolic tetrahedron with maximal volume.

13.1.1. The Lobachevsky function. The *Lobachevsky function* is

$$\Lambda(\theta) = -\int_0^\theta \log|2\sin t|dt.$$

The function $\log|2\sin t|$ is $-\infty$ on $\pi\mathbb{Z}$ but is integrable, hence Λ is well-defined and continuous on $\mathbb{R}$. Its first derivatives are

$$\Lambda'(\theta) = -\log|2\sin\theta|, \qquad \Lambda''(\theta) = -\cot\theta.$$

The function Λ has derivative $+\infty$ on $\pi\mathbb{Z}$ and is an odd function, because its derivative is even.

Proposition 13.1.1. *The function Λ is π-periodic. We have $\Lambda(0) = \Lambda\left(\frac{\pi}{2}\right) = \Lambda(\pi) = 0$. The function Λ is strictly positive on $\left(0, \frac{\pi}{2}\right)$, strictly negative on $\left(\frac{\pi}{2}, \pi\right)$, and has absolute maximum and minimum at $\frac{\pi}{6}$ and $\frac{5}{6}\pi$. For all $m \in \mathbb{N}$ the following holds:*

$$\Lambda(m\theta) = m \sum_{k=0}^{m-1} \Lambda\left(\theta + \frac{k\pi}{m}\right).$$

Proof. We prove the equality for $m = 2$:

$$\frac{\Lambda(2\theta)}{2} = -\frac{1}{2}\int_0^{2\theta} \log|2\sin t|dt = -\int_0^\theta \log|2\sin 2t|dt$$

$$= -\int_0^\theta \log|2\sin t|dt - \int_0^\theta \log\left|2\sin\left(t + \frac{\pi}{2}\right)\right|dt$$

$$= \Lambda(\theta) - \int_{\frac{\pi}{2}}^{\frac{\pi}{2}+\theta} \log|2\sin t|dt$$

$$= \Lambda(\theta) + \Lambda\left(\theta + \frac{\pi}{2}\right) - \Lambda\left(\frac{\pi}{2}\right).$$

By setting $\theta = \frac{\pi}{2}$ we get $\Lambda(\pi) = 0$. Since the derivative Λ' is π-periodic and $\Lambda(\pi) = 0$, also Λ is π-periodic. Since Λ is π-periodic and odd, we have $\Lambda\left(\frac{\pi}{2}\right) = 0$. We have also proved the formula for $m = 2$.

To prove the formula for generic m we use a generalisation of the duplication formula for the sinus. From the equality

$$z^m - 1 = \prod_{k=0}^{m-1}\left(z - e^{-\frac{2\pi i k}{m}}\right)$$

we deduce

$$2\sin(mt) = \prod_{k=0}^{m-1} 2\sin\left(t + \frac{k\pi}{m}\right)$$

and hence

$$\frac{\Lambda(m\theta)}{m} = -\frac{1}{m}\int_0^{m\theta} \log|2\sin t|dt = -\int_0^\theta \log|2\sin(mt)|dt$$

$$= -\sum_{k=0}^{m-1}\int_0^\theta \log\left|2\sin\left(t + \frac{k\pi}{m}\right)\right|dt$$

$$= -\sum_{k=0}^{m-1}\left(\int_0^{\theta+\frac{k\pi}{m}} \log|2\sin t|dt - \int_0^{\frac{k\pi}{m}} \log|2\sin t|dt\right)$$

$$= \sum_{k=0}^{m-1} \Lambda\left(\theta + \frac{k\pi}{m}\right) + C(m)$$
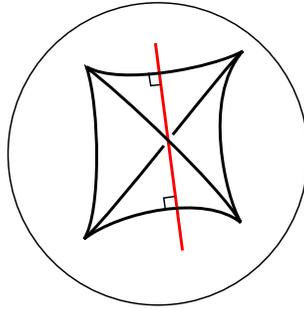
Figure 13.1. Every pair of opposite edges in an ideal tetrahedron has an axis orthogonal to both which is a symmetry axis for the tetrahedron.

where $C(m)$ is a constant independent of θ. By integrating both sides we get

$$\frac{1}{m} \int_0^\pi \Lambda(m\theta) = \sum_{k=0}^{m-1} \int_0^\pi \Lambda\left(\theta + \frac{k\pi}{m}\right) + C(m)\pi.$$

Since Λ is odd and π-periodic, we have

$$\int_0^\pi \Lambda(m\theta) = 0$$

for any integer m. Hence $C(m) = 0$ and the formula is proved. Finally we note that $\Lambda''(\theta) = -\cot\theta$ is strictly negative in $(0, \frac{\pi}{2})$ and strictly positive in $(\frac{\pi}{2}, \pi)$, hence Λ is strictly positive in $(0, \frac{\pi}{2})$ and strictly negative in $(\frac{\pi}{2}, \pi)$. $\qquad\square$

13.1.2. Volumes of ideal tetrahedra. An ideal tetrahedron in $\mathbb{H}^3$ is the convex hull of four non-planar ideal points. Quite surprisingly, every ideal tetrahedron has some non-trivial symmetries.

Proposition 13.1.2. *For any pair of opposite edges in an ideal tetrahedron Δ there is a unique line r orthogonal to both as in Figure 13.1 and Δ is symmetric with respect to a π-rotation around r.*

Proof. The opposite edges e ed e' are ultraparallel lines in $\mathbb{H}^3$ and hence have a common perpendicular r. A π-rotation around r inverts both e and e' but preserve the 4 ideal vertices of Δ, hence Δ itself. $\qquad\square$

As a consequence, two opposite edges in Δ have coinciding dihedral angles as in Figure 13.2-(left). Moreover, we have $\alpha + \beta + \gamma = \pi$ because a small horosphere based at a vertex intersects Δ into a Euclidean triangle with inner angles α, β, and γ. The regular ideal tetrahedron has of course equal angles $\alpha = \beta = \gamma = \frac{\pi}{3}$.
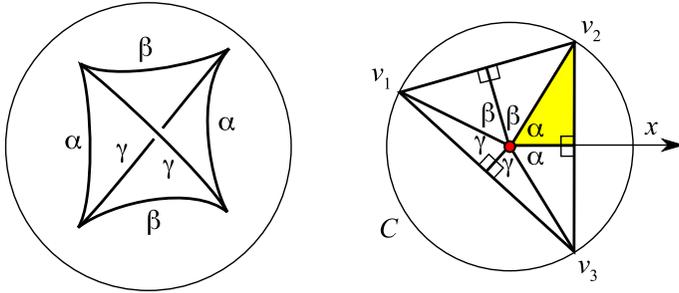
Figure 13.2. The dihedral angles α, β, γ of an ideal tetrahe-
dron. Opposite edges have the same angle and $\alpha + \beta + \gamma = \pi$
(left). To calculate the volume we use the half-space model,
send a vertex to ∞, and divide the tetrahedron in six sub-
tetrahedra (right).

Theorem 13.1.3. *Let Δ be an ideal tetrahedron with dihedral angles
α, β and γ. We have*

$$\mathrm{Vol}(\Delta) = \Lambda(\alpha) + \Lambda(\beta) + \Lambda(\gamma).$$

Proof. We represent Δ in the half-space model H^3 with one vertex v_0
at infinity and three vertices v_1, v_2, v_3 in $\mathbb{C}$. Let C be the circle containing
v_1, v_2, and v_3: up to composing with elements in $\mathbb{PSL}_2(\mathbb{C})$ we can suppose
that $C = S^1$. The Euclidean triangle $T \subset \mathbb{C}$ with vertices v_1, v_2, and v_3
has interior angles α, β, and γ.

We first consider the case $0 \in T$, that is $\alpha, \beta, \gamma \leqslant \frac{\pi}{2}$. We decompose
T into six triangles as in Figure 13.2: the tetrahedron Δ decomposes
accordingly into six tetrahedra lying above them, and we prove that the
one Δ_α lying above the yellow triangle has volume $\frac{\Lambda(\alpha)}{2}$. This proves the
theorem.

The tetrahedron Δ_α is the intersection of four half-spaces: three ver-
tical ones bounded by the hyperplanes $y = 0$, $x = \cos \alpha$, and $y = x \tan \alpha$,
and one bounded by the half-sphere $z^2 = x^2 + y^2$. Therefore

$$\mathrm{Vol}(\Delta_\alpha) = \int_0^{\cos \alpha} dx \int_0^{x \tan \alpha} dy \int_{\sqrt{1-x^2-y^2}}^{\infty} \frac{1}{z^3} dz$$

$$= \int_0^{\cos \alpha} dx \int_0^{x \tan \alpha} dy \left[-\frac{1}{2z^2} \right]_{\sqrt{1-x^2-y^2}}^{\infty}$$

$$= \frac{1}{2} \int_0^{\cos \alpha} dx \int_0^{x \tan \alpha} \frac{1}{1-x^2-y^2} dy.$$

To solve this integral we use the relation

$$\frac{1}{1-x^2-y^2} = \frac{1}{2\sqrt{1-x^2}} \left(\frac{1}{\sqrt{1-x^2}-y} + \frac{1}{\sqrt{1-x^2}+y} \right)$$

and hence $\text{Vol}(\Delta_\alpha)$ equals

$$\frac{1}{4} \int_0^{\cos\alpha} \frac{dx}{\sqrt{1-x^2}} \left(\left[-\log(\sqrt{1-x^2}-y) \right]_0^{x\tan\alpha} + \left[\log(\sqrt{1-x^2}+y) \right]_0^{x\tan\alpha} \right)$$

$$= \frac{1}{4} \int_0^{\cos\alpha} \frac{dx}{\sqrt{1-x^2}} \left(-\log(\sqrt{1-x^2}-x\tan\alpha) + \log(\sqrt{1-x^2}+x\tan\alpha) \right).$$

By writing $x = \cos t$ and hence $dx = -\sin t\, dt$ we obtain

$$\text{Vol}(\Delta_\alpha) = \frac{1}{4} \int_{\frac{\pi}{2}}^{\alpha} \frac{-\sin t}{\sin t} \left(-\log \frac{\sin t \cos\alpha - \cos t \sin\alpha}{\sin t \cos\alpha + \cos t \sin\alpha} \right) dt$$

$$= -\frac{1}{4} \int_{\frac{\pi}{2}}^{\alpha} \log \frac{\sin(t+\alpha)}{\sin(t-\alpha)} dt = -\frac{1}{4} \int_{\frac{\pi}{2}}^{\alpha} \log \frac{|2\sin(t+\alpha)|}{|2\sin(t-\alpha)|} dt$$

$$= \frac{1}{4} \int_{2\alpha}^{\frac{\pi}{2}+\alpha} \log|2\sin t|dt - \frac{1}{4} \int_0^{\frac{\pi}{2}-\alpha} \log|2\sin t|dt$$

$$= \frac{1}{4} \left(-\Lambda\left(\frac{\pi}{2}+\alpha\right) + \Lambda(2\alpha) + \Lambda\left(\frac{\pi}{2}-\alpha\right) \right)$$

$$= \frac{1}{4} \left(-\Lambda\left(\frac{\pi}{2}+\alpha\right) + 2\Lambda(\alpha) + 2\Lambda\left(\frac{\pi}{2}+\alpha\right) - \Lambda\left(\frac{\pi}{2}+\alpha\right) \right)$$

$$= \frac{1}{2}\Lambda(\alpha)$$

using Proposition 13.1.1.

If $0 \notin T$ the triangle T may be decomposed analogously into triangles, some of which contribute negatively to the volume, and we obtain the same formula.   □

Corollary 13.1.4. *The regular ideal tetrahedron is the hyperbolic tetrahedron of maximum volume.*

Proof. It is easy to prove that every hyperbolic tetrahedron is contained in an ideal tetrahedron: hence we may consider only ideal tetrahedra. Consider the triangle $T = \{0 \leqslant \alpha, \beta, \alpha+\beta \leqslant \pi\}$ and

$$f: \quad T \longrightarrow \mathbb{R}$$
$$(\alpha, \beta) \longmapsto \Lambda(\alpha) + \Lambda(\beta) + \Lambda(\pi - \alpha - \beta).$$

The continuous function f is null on ∂T and strictly positive on the interior of T because it measures the volume of the ideal tetrahedron of dihedral angles $\alpha, \beta, \gamma = \pi - \alpha - \beta$. Hence f has at least a maximum on some interior point (α, β). The gradient ∇f is equal to

$$\begin{pmatrix} \Lambda'(\alpha) - \Lambda'(\pi - \alpha - \beta) \\ \Lambda'(\beta) - \Lambda'(\pi - \alpha - \beta) \end{pmatrix} = \begin{pmatrix} -\log|2\sin\alpha| + \log|2\sin(\pi - \alpha - \beta)| \\ -\log|2\sin\beta| + \log|2\sin(\pi - \alpha - \beta)| \end{pmatrix}$$

and it must vanish there: this holds if and only if $\sin\alpha = \sin(\pi - \alpha - \beta) = \sin\beta$, *i.e.* if and only if the tetrahedron has all dihedral angles $\frac{\pi}{3}$.   □

13.2. Simplicial volume

Gromov has introduced a topological invariant on closed manifolds of any dimension called the *simplicial volume*. This nice invariant can be used, among other things, to prove Mostow's rigidity theorem.

13.2.1. Definition. Gromov has introduced a measure of "volume" of a closed manifold M which makes use only of the homology of M. Quite surprisingly, this notion of volume coincides (up to a factor) with the Riemannian one when M is hyperbolic.

Consider a topological space X and its homology with ring $\mathbb{R}$. We define the *norm* of a cycle $\alpha = \lambda_1\alpha_1 + \ldots + \lambda_h\alpha_h$ as follows:

$$|\alpha| = |\lambda_1| + \ldots + |\lambda_h|.$$

Definition 13.2.1. The *norm* of a class $a \in H_k(X, \mathbb{R})$ is the infimum of the norms of its elements:

$$|a| = \inf\big\{|\alpha| \ \big| \ \alpha \in Z_k(X, \mathbb{R}), [\alpha] = a\big\}.$$

Recall that a *seminorm* on a real vector space V is a map $|\cdot| \colon V \to \mathbb{R}_{\geqslant 0}$ such that

- $|\lambda v| = |\lambda||v|$ for any scalar $\lambda \in \mathbb{R}$ and vector $v \in V$,
- $|v + w| \leqslant |v| + |w|$ for any pair of vectors $v, w \in V$.

A *norm* is a seminorm where $|v| = 0$ implies $v = 0$. The following is immediate.

Proposition 13.2.2. *The norm* $|\cdot|$ *induces a seminorm on* $H_k(X, \mathbb{R})$.

Although it is only a seminorm, the function $|\cdot|$ is called a norm for simplicity. Let now M be an oriented closed connected manifold: we know that $H_n(M, \mathbb{Z}) \cong \mathbb{Z}$ and the orientation of M determines a *fundamental class* $[M] \in H_n(M, \mathbb{Z})$ that generates the group. Moreover $H_n(M, \mathbb{R}) \cong \mathbb{R}$ and there is a natural inclusion

$$\mathbb{Z} \cong H_n(M, \mathbb{Z}) \hookrightarrow H_n(M, \mathbb{R}) \cong \mathbb{R}$$

hence the fundamental class $[M]$ is naturally in $H_n(M, \mathbb{R})$ and has a norm.

Definition 13.2.3. The *simplicial volume* $\|M\| \in \mathbb{R}_{\geqslant 0}$ of a closed oriented connected M is the norm of its fundamental class:

$$\|M\| = |[M]|$$

Since $|[M]| = |-[M]|$ the simplicial volume actually does not depend on the orientation. When M is non-orientable we set $\|M\| = \|\tilde{M}\|/2$ where $\tilde{M}$ is the orientable double cover of M. The definition of $\|M\|$ is relatively simple but has various non-obvious consequences.

13.2.2. Properties. A continuous map $f\colon M \to N$ between closed oriented n-manifolds induces a homomorphism $f_*\colon H_n(M, \mathbb{Z}) \to H_n(N, \mathbb{Z})$, and recall that the *degree* of f is the integer $\deg f$ such that

$$f_*([M]) = \deg f \cdot [N].$$

Proposition 13.2.4. *Let* $f\colon M \to N$ *be a continuous map between closed oriented manifolds. The following inequality holds:*

$$\|M\| \geqslant |\deg f| \cdot \|N\|.$$

Proof. Every description of $[M]$ as a cycle $\lambda_1\alpha_1 + \ldots + \lambda_h\alpha_h$ induces a description of $f_*([M]) = \deg f[N]$ as a cycle $\lambda_1 f \circ \alpha_1 + \ldots + \lambda_h f \circ \alpha_h$ with the same norm (or less, if there is some cancelation). $\square$

Corollary 13.2.5. *If M and N are closed orientable and homotopically equivalent n-manifolds then* $\|M\| = \|N\|$.

Proof. A homotopic equivalence consists of two maps $f\colon M \to N$ and $g\colon N \to M$ whose compositions are both homotopic to the identity. In particular both f and g have degree ± 1. $\square$

Corollary 13.2.6. *If M admits a continuous self-map* $f\colon M \to M$ *of degree at least two then* $\|M\| = 0$.

Corollary 13.2.7. *Every sphere* S^n *has norm zero. More generally we have* $\|M \times S^n\| = 0$ *for every closed M and any* $n \geqslant 1$.

Proof. A sphere S^n admits self-maps of degree $\geqslant 2$, and hence also the product $M \times S^n$ does. $\square$

Among the genus-g surfaces S_g, we deduce that the sphere and the torus have simplicial volume zero. We will see soon that every surface of genus $g \geqslant 2$ has positive simplicial volume. When the continuous map is a covering the inequality from Proposition 13.2.4 is promoted to an equality.

Proposition 13.2.8. *If* $f\colon M \to N$ *is a degree-d covering we have*

$$\|M\| = d \cdot \|N\|.$$

Proof. The reason for this equality is that cycles can be lifted and projected through the covering. More precisely, we already know that $\|M\| \geqslant d \cdot \|N\|$. Conversely, let $\alpha = \lambda_1\alpha_1 + \ldots + \lambda_h\alpha_h$ represent $[N]$; each α_i is a map $\Delta_n \to N$. Since Δ_n is simply connected, the map α_i lifts to d distinct maps $\alpha_i^1, \ldots, \alpha_i^d\colon \Delta_n \to N$. The chain $\tilde{\alpha} = \sum_{ij} \lambda_i\alpha_i^j$ is a cycle in M and $f_*(\tilde{\alpha}) = d\alpha$. Hence $\|M\| \leqslant d \cdot \|N\|$. $\square$

We also note the following fact.

Proposition 13.2.9. *If M is triangulated with k simplices,* $\|M\| \leqslant k$.

Proof. Up to taking a double cover we may suppose that M is oriented. The closed n-manifold M is triangulated into simplices $\Delta_1, \ldots, \Delta_k$, and we fix an orientation-preserving parametrisation $s_i\colon \Delta \to \Delta_i$ of each. We would like to say that $s_1 + \ldots + s_k$ is a fundamental cycle, however this singular chain is not necessarily a cycle because the restriction to adjacent faces coincide only up to the symmetries S_{n+1} of Δ.

We can fix this problem easily by averaging each s_i on all its permutations, that is we substitute each s_i with $\frac{1}{(n+1)!} \sum_{\sigma \in S_{n+1}} (-1)^{\mathrm{sgn}(\sigma)} s_i \circ \sigma$. Now $s = s_1 + \ldots + s_k$ is a fundamental cycle and $|s| = k$. $\qquad\square$

13.2.3. Seifert manifolds. We can now calculate the simplicial volume of Seifert manifolds.

Proposition 13.2.10. *If M is a closed Seifert 3-manifold then $\|M\| = 0$.*

Proof. Every Seifert manifold is finitely covered by a product $S \times S^1$ (if $e = 0$) or by a bundle $(S, (1, 1))$ with Euler number 1 (if $e \neq 0$). In the first case we are done since $\|S \times S^1\| = 0$.

In the second, if $S = S^2$ then $(S^2, (1, 1)) = S^3$ and $\|S^3\| = 0$, so we suppose $\chi(S) \leqslant 0$. There is a universal $K > 0$ such that $(S, (1, 1))$ triangulates with at most $K|\chi(S)| + K$ tetrahedra (exercise) and therefore $\|(S, (1, 1))\| \leqslant K|\chi(S)| + K$. Exercise 10.3.27 shows that for every $e > 0$ there is a degree-e covering $\tilde{S} \to S$ and two degree-e coverings

$$(\tilde{S}, (1, 1)) \longrightarrow (\tilde{S}, (1, e)) \longrightarrow (S, (1, 1))$$

which compose to a degree-e^2 covering $(\tilde{S}, (1, 1)) \to (S, (1, 1))$. Thus

$$\|(S, (1, 1))\| = \frac{\|(\tilde{S}, (1, 1))\|}{e^2} \leqslant \frac{K|\chi(\tilde{S})| + K}{e^2} \leqslant \frac{Ke|\chi(S)| + K}{e^2} \to 0$$

as $e \to \infty$. $\qquad\square$

13.2.4. Simplicial and hyperbolic volume. In the next pages we will prove the following theorem. Let v_3 be the volume of the regular ideal tetrahedron in $\mathbb{H}^3$.

Theorem 13.2.11. *Let M be a closed hyperbolic 3-manifold. We have*

$$\mathrm{Vol}(M) = v_3 \|M\|.$$

The theorem furnishes in particular some examples of manifolds with positive simplicial volume and shows that $\mathrm{Vol}(M)$ is a topological invariant of M, thus generalising the Gauss-Bonnet theorem to dimension $n = 3$. Mostow rigidity will then strengthen this result in dimension $n = 3$, showing that the hyperbolic metric itself (not only its volume) is a topological invariant.

Both quantities $\mathrm{Vol}(M)$ and $\|M\|$ are multiplied by d if we substitute M with a degree-d covering. In particular, up to substituting M with its orientable double cover we can suppose that M is orientable.

13.2.5. Cycle straightening. The *straight singular k-simplex* with vertices $v_0, \ldots, v_k \in \mathbb{H}^n$ is the map

$$\alpha: \quad \Delta_k \longrightarrow \mathbb{H}^n$$
$$(t_0, \ldots, t_k) \longmapsto t_0 v_0 + \ldots + t_k v_k$$

defined using convex combinations, see Section 2.4.3. If the $k + 1$ vertices are not contained in a $(k - 1)$-plane the singular k-simplex is *non-degenerate* and its image is a hyperbolic k-simplex.

The *straightening* α^{st} of a singular simplex $\alpha: \Delta_k \to \mathbb{H}^n$ is the straight singular simplex with the same vertices of α. The straightening α^{st} of a singular simplex $\alpha: \Delta_k \to M$ in a hyperbolic manifold $M = \mathbb{H}^n/_\Gamma$ is defined by lifting the singular simplex in $\mathbb{H}^n$, straightening it, and projecting it back to M by composition with the covering map. Different lifts produce the same straightening in M because they are related by isometries of $\mathbb{H}^n$.

The straightening extends by linearity to a homomorphism

$$\mathrm{st}: C_k(M, R) \to C_k(M, R)$$

which commutes with ∂ and hence induces a homomorphism in homology

$$\mathrm{st}_*: H_k(M, R) \to H_k(M, R).$$

Proposition 13.2.12. *The map* st_* *is the identity.*

Proof. We may define a homotopy between a singular simplex σ and its straightening σ^{st} using the convex combination

$$\sigma^t(x) = t\sigma(x) + (1 - t)\sigma^{\mathrm{st}}(x).$$

This defines a chain homotopy between st_* and id via the same technique used to prove that homotopic maps induce the same maps in homology. $\square$

The *abstract volume* of a straightened singular n-simplex $\alpha: \Delta_n \to M$ is the volume of its lift in $\mathbb{H}^n$ and may also be calculated as

$$\left| \int_\alpha \omega \right|$$

where ω is the volume form on M pulled back along α. If α is non-degenerate, we say that its *sign* is *positive* if α is orientation-preserving and *negative* otherwise: equivalently, it is the sign of $\int_\alpha \omega$.

We now concentrate on the dimension $n = 3$ where we know that a tetrahedron has maximum volume v_3 if and only if it is both regular and ideal: a straight simplex is compact and hence its abstract volume is strictly smaller than v_3. We can now easily prove one inequality.

Proposition 13.2.13. *Let M be a closed hyperbolic 3-manifold. We have*

$$\mathrm{Vol}(M) \leqslant v_3 \|M\|.$$

Proof. As we said above, we can suppose that M is orientable. Take a cycle $\alpha = \lambda_1\alpha_1 + \ldots + \lambda_k\alpha_k$ that represents $[M]$. We can suppose it is straightened, because the straightening preserves both the coefficients and the homology class. Let ω be the volume form on M. We get

$$\mathrm{Vol}(M) = \int_M \omega = \int_\alpha \omega = \lambda_1 \int_{\alpha_1} \omega + \ldots + \lambda_k \int_{\alpha_k} \omega.$$

The quantity $\left| \int_{\alpha_i} \omega \right|$ is the abstract volume of α_i. Hence $\left| \int_{\alpha_i} \omega \right| < v_3$ and

$$\mathrm{Vol}(M) < \big(|\lambda_1| + \ldots + |\lambda_k|\big) v_3.$$

This holds for all α, hence $\mathrm{Vol}(M) \leqslant v_3 \|M\|$. $\qquad\square$

The proof of the converse inequality is less immediate.

13.2.6. Efficient cycles. Let $M = \mathbb{H}^3/_\Gamma$ be a closed oriented hyperbolic 3-manifold. An ε-*efficient cycle* for M is a straightened cycle

$$\alpha = \lambda_1\alpha_1 + \ldots + \lambda_k\alpha_k$$

representing $[M]$ where the abstract volume of α_i if bigger than $v_3 - \varepsilon$ and the sign of α_i is coherent with the sign of λ_i, for all i.

We will construct an ε-efficient cycle for every $\varepsilon > 0$. This will conclude the proof of Theorem 13.2.11 in virtue of the following:

Lemma 13.2.14. *If for every $\varepsilon > 0$ the manifold M admits an ε-efficient cycle, then we have* $\mathrm{Vol}(M) \geqslant v_3 \|M\|$.

Proof. Let $\alpha = \lambda_1\alpha_1 + \ldots \lambda_k\alpha_k$ be an ε-efficient cycle and ω be the volume form on M. Coherence of signs gives $\lambda_i \int_{\alpha_i} \omega > 0$ for all i. We get

$$\mathrm{Vol}(M) = \int_M \omega = \int_\alpha \omega = \lambda_1 \int_{\alpha_1} \omega + \ldots + \lambda_k \int_{\alpha_k} \omega$$
$$\geqslant \big(|\lambda_1| + \ldots + |\lambda_k|\big) \cdot (v_3 - \varepsilon).$$

Therefore $\mathrm{Vol}(M) \geqslant \|M\| \cdot (v_3 - \varepsilon)$ for all $\varepsilon > 0$. $\qquad\square$

It remains to construct ε-efficient cycles.

Exercise 13.2.15. If Δ^i is a sequence of tetrahedra in $\mathbb{H}^3$ whose vertices tend to the vertices of a regular ideal tetrahedron in $\partial\mathbb{H}^3$, then

$$\mathrm{Vol}(\Delta^i) \to v_3.$$

For any $t > 0$, let $\Delta(t)$ be a regular tetrahedron obtained as in Section 3.2.4 as follows. Pick a point $x \in \mathbb{H}^3$ and a regular tetrahedron in the Euclidean tangent space $T_x\mathbb{H}^3$, centred at the origin with vertices at distance t from it, project the vertices in $\mathbb{H}^3$ via the exponential map, and pick their convex hull.

A t-*simplex* is a tetrahedron isometric to $\Delta(t)$ equipped with an ordering of its vertices: the ordering allows us to consider it as a straightened singular simplex. Let $S(t)$ be the set of all t-simplices in $\mathbb{H}^3$.

Exercise 13.2.16. The group $\mathrm{Isom}(\mathbb{H}^3)$ acts on $S(t)$ freely and transitively.

Recall from Corollary 2.4.8 that $\mathrm{Isom}(\mathbb{H}^3)$ is unimodular: the Haar measure on $\mathrm{Isom}(\mathbb{H}^3)$ induces an $\mathrm{Isom}(\mathbb{H}^3)$-invariant measure on $S(t)$.

Let $M = \mathbb{H}^3/_\Gamma$ be a closed hyperbolic 3-manifold and $\pi \colon \mathbb{H}^3 \to M$ the covering projection. Fix a base point $x_0 \in \mathbb{H}^3$ and consider its orbit $O = \Gamma x_0$. Consider the set

$$\Sigma = \Gamma^4/_\Gamma$$

of the 4-uples (g_0, g_1, g_2, g_3) considered up to the diagonal action of Γ:

$$g \cdot (g_0, g_1, g_2, g_3) = (gg_0, gg_1, gg_2, gg_3).$$

An element $\sigma = (g_0, g_1, g_2, g_3) \in \Sigma$ determines a singular simplex $\tilde{\sigma}$ in $\mathbb{H}^3$ with vertices $g_0(x_0), g_1(x_0), g_2(x_0), g_3(x_0) \in O$ only up to translations by $g \in \Gamma$, hence it gives a well-defined singular simplex in M, which we still denote by σ. We now introduce the chain

$$\alpha(t) = \sum_{\sigma \in \Sigma} \lambda_\sigma(t) \cdot \sigma$$

for some appropriate real coefficients $\lambda_\sigma(t)$ that we now define. The base point x_0 determines the Dirichlet tessellation of $\mathbb{H}^3$ into domains $D(g(x_0))$, $g \in \Gamma$. For $\sigma = (g_0, g_1, g_2, g_3)$ we let $S_\sigma^+(t) \subset S(t)$ be the set of all positive t-simplices whose i-th vertex lies in $D(g_i(x_0))$ for all i. The number $\lambda_\sigma^+(t)$ is the measure of $S_\sigma^+(t)$. We define analogously $\lambda_\sigma^-(t)$ and set

$$\lambda_\sigma(t) = \lambda_\sigma^+(t) - \lambda_\sigma^-(t).$$

Lemma 13.2.17. *The chain $\alpha(t)$ has finitely many addenda and is a cycle. If t is sufficiently big the cycle $\alpha(t)$ represents a positive multiple of $[M]$ in the group $H_3(M, \mathbb{R})$.*

Proof. We prove that the sum is finite. Let d, T be the diameters of $D(x_0)$ and of a t-simplex. We write $\sigma = (\mathrm{id}, g_1, g_2, g_3)$ for all $\sigma \in \Sigma$: that is, all simplices have their first vertex at x_0. If $\lambda_\sigma(t) \neq 0$ then $d(g_i x_0, x_0) < 2d + T$ for all i: therefore $\alpha(t)$ has finitely many addenda (because O is discrete).

We prove that $\alpha(t)$ is a cycle. The boundary $\partial\alpha(t)$ is a linear combination of straight 2-simplices with vertices in $(g_0 x_0, g_1 x_0, g_2 x_0)$ as g_0, g_1, and g_2 vary. The coefficient of one such 2-simplex is

$$\sum_{g \in \Gamma} \big(-\lambda_{(g,g_0,g_1,g_2)}(t) + \lambda_{(g_0,g,g_1,g_2)}(t) - \lambda_{(g_0,g_1,g,g_2)}(t) + \lambda_{(g_0,g_1,g_2,g)}(t) \big).$$

We prove that each addendum summed along $g \in \Gamma$ is zero; for simplicity we consider the last addendum and get

$$\sum_{g \in \Gamma} \lambda_{(g_0,g_1,g_2,g)}(t) = \sum_{g \in \Gamma} \lambda_{(g_0,g_1,g_2,g)}(t)^+ - \sum_{g \in \Gamma} \lambda_{(g_0,g_1,g_2,g)}(t)^-.$$

The first addendum measures the positive t-simplices whose first 3 vertices lie in $D(g_0(x_0)), \dots, D(g_2(x_0))$, the second measures the negative t-simplices with the same requirement. These two subsets have the same volume in $S(t)$ because they are related by the involution $r\colon S(t) \to S(t)$ that mirrors a simplex with respect to its first facet.

We show that for sufficiently big t the cycle is a positive multiple of $[M]$. Let t be sufficiently big so that two vertices in a t-simplex have distance bigger than $2d$. This condition implies that if there is a positive t-simplex with vertices in $D(g_0(x_0)), \dots, D(g_3(x_0))$, then any straight simplex with vertices in $D(g_0(x_0)), \dots, D(g_3(x_0))$ is positive. Therefore in the expression

$$\alpha(t) = \sum_{\sigma \in \Sigma} \lambda_\sigma(t) \cdot \sigma$$

the signs of $\lambda_\sigma(t)$ and σ are coherent and

$$\int_{\alpha(t)} \omega = \sum_{\sigma \in \Sigma} \lambda_\sigma(t) \cdot \int_\sigma \omega > 0.$$

Therefore $\alpha(t)$ is a positive multiple of $[M]$. $\qquad\qquad\square$

For sufficiently big t we have $\alpha(t) = k_t[M]$ in homology for some $k_t > 0$. The rescaled $\bar\alpha(t) = \alpha(t)/k_t$ hence represents $[M]$. We have found our ε-efficient cycles.

Lemma 13.2.18. *For any $\varepsilon > 0$ there is a $t_0 > 0$ such that $\bar\alpha(t)$ is ε-efficient for all $t > t_0$.*

Proof. Let d be the diameter of the Dirichlet domain $D(x_0)$. Let a *quasi t-simplex* be a simplex whose vertices are at distance $< d$ from those of a t-simplex. By construction $\bar\alpha(t)$ is a linear combination of quasi t-simplices.

We now show that for any $\varepsilon > 0$ there is a $t_0 > 0$ such that for all $t > t_0$ every quasi t-simplex has volume bigger than $v_3 - \varepsilon$. By contradiction, let Δ^t be a sequence of quasi t-simplices of volume smaller than $v_3 - \varepsilon$ with $t \to \infty$. The vertices of Δ^t are d-closed to a t-simplex Δ^t_*, and we move the pair Δ^t, Δ^t_* isometrically so that the t-simplices Δ^t_* have the same barycenter. Now both the vertices of Δ^t and Δ^t_* tend to the vertices of an ideal regular tetrahedron and Exercise 13.2.15 gives a contradiction. $\quad\square$

The previous lemmas together prove the second half of Theorem 13.2.11.

Corollary 13.2.19. *Let M be a closed hyperbolic 3-manifold. We have*

$$\mathrm{Vol}(M) \geqslant v_3\|M\|.$$

Theorem 13.2.11 has some non-trivial consequences.

Corollary 13.2.20. *Let M, N be two closed orientable hyperbolic 3-manifolds. If there is a map $f \colon M \to N$ of degree d then $\mathrm{Vol}(M) \geqslant |d| \cdot \mathrm{Vol}(N)$.*

Corollary 13.2.21. *Two homotopically equivalent and closed hyperbolic 3-manifolds have the same volume.*

Remark 13.2.22. If we are able to prove that the regular ideal simplex has the maximum volume v_n among hyperbolic n-simplices in $\mathbb{H}^n$, then the whole proof extends as is from $\mathbb{H}^3$ to $\mathbb{H}^n$ and shows that $\mathrm{Vol}(M) = v_n \cdot |M|$ for any closed hyperbolic n-manifold M. This is obviously true when $n = 2$, and we get $v_2 = \pi$.

13.3. Mostow rigidity

13.3.1. Introduction.
We want to prove the following.

Theorem 13.3.1 (Mostow rigidity). *Let M and N be closed connected orientable hyperbolic 3-manifolds. Every isomorphism $\pi_1(M) \xrightarrow{\sim} \pi_1(N)$ between fundamental groups is induced by a unique isometry $M \xrightarrow{\sim} N$.*

This powerful theorem says that an algebraic isomorphism between fundamental groups suffices to produce and characterise an isometry.

Corollary 13.3.2. *Two closed orientable hyperbolic 3-manifolds with isomorphic fundamental groups are isometric.*

We prove Mostow's rigidity in this section. All the ingredients are already there, we only need to make a little last effort.

We note that closed hyperbolic manifolds are aspherical because their universal cover $\mathbb{H}^n$ is contractible. For such manifolds every isomorphism $\pi_1(M) \to \pi_1(N)$ is induced by a homotopy equivalence $f \colon M \to N$, unique up to homotopy: see Corollary 1.7.7. To prove Mostow's theorem we need to promote this homotopy equivalence to an isometry in dimension $n = 3$. We already know that $\mathrm{Vol}(M) = \mathrm{Vol}(N)$ by Corollary 13.2.21.

13.3.2. Proof of Mostow's theorem.
Let $f \colon M \to N$ be a smooth homotopy equivalence. Recall from Theorem 5.2.1 that f lifts to a map $\tilde{f} \colon \mathbb{H}^3 \to \mathbb{H}^3$ which extends continuously to a homeomorphism $\tilde{f} \colon \partial \mathbb{H}^3 \to \partial \mathbb{H}^3$ of the boundary spheres. We start with a lemma.

Lemma 13.3.3. *The extension $\tilde{f} \colon \partial \mathbb{H}^3 \to \partial \mathbb{H}^3$ sends the vertices of every regular ideal simplex to the vertices of some regular ideal simplex.*

Proof. Let $w_0, \dots, w_3$ be vertices of a regular ideal simplex and suppose by contradiction that their images $\tilde{f}(w_0), \dots, \tilde{f}(w_3)$ span a non-regular ideal simplex, which has volume smaller than $v_3 - 2\delta$ for some $\delta > 0$. By Exercise 13.2.15 there are neighbourhoods U_i of w_i in $\overline{\mathbb{H}^3}$ for $i = 0, \dots, 3$ such that the volume of the simplex with vertices $\tilde{f}(u_0), \dots, \tilde{f}(u_3)$ is smaller than $v_3 - \delta$ for any choice of $u_i \in U_i$.

In Section 13.2.6 we have defined a cycle

$$\alpha(t) = \sum_{\sigma \in \Sigma} \lambda_\sigma(t) \cdot \sigma$$

where t depends on ε. We say that a singular simplex $\sigma \in \Sigma$ is *bad* if its i-th vertex is contained in U_i for all i. Let $\Sigma^{\text{bad}} \subset \Sigma$ be the subset of all bad singular simplices and define

$$\alpha(t)^{\text{bad}} = \sum_{\sigma \in \Sigma^{\text{bad}}} \lambda_\sigma(t) \cdot \sigma.$$

We want to estimate $|\alpha(t)|$ and $|\alpha(t)^{\text{bad}}|$. We prove that

$$|\alpha(t)| = \sum_{\sigma \in \Sigma} |\lambda_\sigma(t)|$$

is a real number independent of t: let $S_0 \subset S(t)$ be the set of all t-simplices having the first vertex in the Dirichlet domain $D(x_0)$ of the fixed base point $x_0 \in \mathbb{H}^3$. It follows from the definitions that $|\alpha(t)|$ equals the measure of S_0 for sufficiently big t. Moreover the set S_0 is in natural correspondence with the set of all isometries that send x_0 to some point in $D(x_0)$: its volume does not depend on t.

To estimate $|\alpha(t)^{\text{bad}}|$ we fix $g_0 \in \Gamma$ so that $D(g_0 x_0) \subset U_0$. Let $S^{\text{bad}} \subset S(t)$ be the set of all bad t-simplices with first vertex in $D(g_0 x_0)$. If t is sufficiently big, the volume of S^{bad} is bigger than a constant independent of t (exercise).

We have proved that $|\alpha(t)^{\text{bad}}|/|\alpha(t)| > C > 0$ independently of t. We may suppose that $\alpha(t)$ represents $[M]$ up to renormalising. The map $f \colon M \to N$ has degree ± 1 and it sends $\alpha(t)$ to a class

$$f_*(\alpha(t)) = \sum_{\sigma \in \Sigma} \lambda_\sigma(t) \cdot (f \circ \sigma)^{\text{st}}$$

representing $\pm[N]$. Since a C-portion of $\alpha(t)$ is bad, a C-portion of simplices in $f_*(\alpha(t))$ has volume smaller than $v_3 - \delta$ and hence

$$\text{Vol}(N) = \left| \int_{f_*(\alpha(t))} \omega \right| < |\alpha(t)|((1 - C)v_3 + C(v_3 - \delta)) = |\alpha(t)|(v_3 - \delta C).$$

Since this holds for all t and $|\alpha(t)| \to \|M\|$ we get

$$\text{Vol}(N) < \|M\|(v_3 - \delta C) = \text{Vol}(M) - \delta C \cdot \|M\|.$$

Corollary 13.2.21 gives $\text{Vol}(M) = \text{Vol}(N)$, a contradiction. $\square$

Proposition 13.3.4. *Every ideal triangle in* $\mathbb{H}^3$ *is the face of precisely two regular ideal tetrahedra.*

Proof. Pick the line l orthogonal to the barycenter of the triangle: the vertex of a regular ideal tetrahedron must be an endpoint of l. To prove that these vertices give regular ideal tetrahedra, note that all ideal triangles in $\mathbb{H}^3$ are isometric, so one concrete example suffices. $\square$

We turn back to Mostow rigidity.

Proposition 13.3.5. *Let $f\colon M \to N$ be a smooth homotopic equivalence between closed hyperbolic orientable 3-manifolds. The restriction $\tilde{f}|_{\partial\mathbb{H}^3}\colon \partial\mathbb{H}^3 \to \partial\mathbb{H}^3$ is the trace of an isometry $\psi\colon \mathbb{H}^3 \to \mathbb{H}^3$.*

Proof. Let $v_0, \ldots, v_3 \in \partial\mathbb{H}^n$ be vertices of a regular ideal tetrahedron Δ. By Lemma 13.3.3 the lift $\tilde{f}$ sends them to the vertices of some regular ideal tetrahedron, and let ψ be the unique isometry of $\mathbb{H}^3$ such that $\psi(v_i) = \tilde{f}(v_i)$ for all i.

By Proposition 13.3.4 there is a unique point $v_4 \neq v_3$ such that v_0, v_1, v_2, v_4 are the vertices of an ideal regular tetrahedron, and $\psi(v_4)$ is the unique point other than $\psi(v_3)$ such that $v_0, v_1, v_2, \psi(v_4)$ are the vertices of an ideal regular tetrahedron. By Lemma 13.3.3 we also have $\tilde{f}(v_4) = \psi(v_4)$.

If we mirror Δ along its faces iteratively we get a tessellation of $\mathbb{H}^3$ via regular ideal tetrahedra, whose ideal vertices form a dense subset of $\partial\mathbb{H}^3$. By iterating this argument in all directions the functions ψ and $\tilde{f}$ coincide on this dense subset and hence on the whole of $\partial\mathbb{H}^3$. $\square$

We can finally prove Mostow's rigidity theorem.

Theorem 13.3.6 (Mostow Rigidity). *Let $f\colon M \to N$ be a homotopic equivalence between closed orientable hyperbolic 3-manifolds. The map f is homotopically equivalent to an isometry.*

Proof. Set $M = \mathbb{H}^3/\Gamma$ and $N = \mathbb{H}^3/\Gamma'$, and pick a lift $\tilde{f}$. We have

$$(18) \qquad \tilde{f} \circ g = f_*(g) \circ \tilde{f} \quad \forall g \in \Gamma$$

for an isomorphism $f_*\colon \Gamma \to \Gamma'$. We may suppose f smooth. The boundary extension of $\tilde{f}$ is the trace of an isometry $\psi\colon \mathbb{H}^3 \to \mathbb{H}^3$ and hence

$$(19) \qquad \psi \circ g = f_*(g) \circ \psi \quad \forall g \in \Gamma$$

holds at $\partial\mathbb{H}^3$. Both terms in (19) are isometries, and isometries are determined by their boundary traces: hence (19) holds for all points in $\mathbb{H}^3$. Therefore ψ descends to an isometry

$$\psi\colon M \to N.$$

A homotopy between f and ψ may be constructed from a convex combination of $\tilde{f}$ and ψ in $\mathbb{H}^n$, which is also Γ-equivariant and hence descends. $\square$

13.3.3. Consequences of Mostow rigidity. The most important consequence is that the entire geometry of a closed hyperbolic 3-manifold is a topological invariant: numerical quantities like the volume of the manifold, its geodesic spectrum, etc. depend only on the topology of the manifold. We single out another application.

Corollary 13.3.7. *Let M be a closed orientable hyperbolic 3-manifold. The natural map*

$$\mathrm{Isom}(M) \to \mathrm{Out}(\pi_1(M))$$

is an isomorphism.

Proof. We already know that it is injective by Proposition 4.3.6. We prove that it is surjective: every automorphism of $\pi_1(M)$ is represented by a homotopy equivalence since M is aspherical (see Corollary 1.7.7), which is in turn homotopic to an isometry by Mostow's rigidity.   □

We note that this is false in dimension $n = 2$, where $\mathrm{Isom}(S)$ is finite and $\mathrm{Out}(\pi_1(S))$ is infinite.

13.3.4. Orbifolds and cone manifolds. While by Mostow's theorem an orientable closed 3-manifold M can have at most one hyperbolic structure, it may have plenty of hyperbolic cone manifold or orbifold structures, typically distinguished by their singular sets and their cone angles. For instance $M = S^3$ has plenty of such structures, as we will see.

For the moment we note that the volumes of these hyperbolic orbifold structures may be arbitrarily big, but not arbitrarily small:

Proposition 13.3.8. *Let a closed orientable 3-manifold M be the underlying space of a hyperbolic orbifold O. We have*

$$\mathrm{Vol}(O) \geqslant v_3 \|M\|.$$

Proof. By Selberg's Lemma there is a degree-d orbifold covering $N \to O$ which is a closed hyperbolic manifold. We have $\mathrm{Vol}(N) = d\mathrm{Vol}(O)$, and since the covering is a degree-d map $N \to M$ we can apply Corollary 13.2.20 and get $\|M\| \leqslant \frac{\|N\|}{d} = \frac{\mathrm{Vol}(N)}{dv_3} = \frac{\mathrm{Vol}(O)}{v_3}$.   □

13.3.5. References. Most of the proofs presented in this chapter were taken from Benedetti − Petronio [**4**], and are also contained in Thurson's notes [**56**]. The proof of Mostow's rigidity presented here was proposed by Gromov and applies to every dimension $n \geqslant 3$, because ideal regular simplexes have indeed the maximum volume by Haagerup and Munkholm [**24**].

Hyperbolic three-manifolds

We have studied and classified the three-manifolds having seven of the eight geometries, and we are now left with the most interesting ones: hyperbolic three-manifolds.

In dimension two every closed hyperbolic surface is constructed by gluing some geodesic pair-of-pants. In dimension three, although closed hyperbolic three-manifolds are everywhere, it is somehow harder to construct them explicitly: the most general procedure to determine a hyperbolic metric (if any) on a given closed 3-manifold consists of solving *Thurston's equations*.

Thurston's equations arise naturally in the attempt of constructing a hyperbolic three-manifold by triangulating it into hyperbolic tetrahedra. The most relevant and unexpected aspect of the theory is that it is much easier to employ hyperbolic *ideal* tetrahedra than compact ones. The combinatorial framework in the ideal case is so convenient, that we use it also for closed three-manifolds.

In this chapter we describe these equations and use them to determine various finite-volume hyperbolic 3-manifolds. We start with the cusped case and then turn to the slightly more complicate closed one.

14.1. Cusped three-manifolds

We know that every cusped finite-volume complete hyperbolic surface is constructed by gluing isometrically finitely many ideal triangles along their edges (see Proposition 7.4.6). Likewise, we now construct plenty of cusped hyperbolic three-manifolds by gluing isometrically finitely many ideal tetrahedra along their faces.

14.1.1. Ideal tetrahedra. Ideal triangles are all isometric, but ideal tetrahedra are not! We now show that they can be described up to isometry by a single complex parameter z with $\Im z > 0$.

An ideal tetrahedron is determined by its ideal vertices $v_1, v_2, v_3, v_4 \in \partial \mathbb{H}^3$. We use the half-space model H^3 and recall that $\partial H^3 = \mathbb{C} \cup \{\infty\}$ and $\mathrm{Isom}^+(H^3) = \mathbb{PSL}_2(\mathbb{C})$, hence there is a unique orientation-preserving isometry of H^3 that sends the vertices v_1, v_2, v_3, v_4 respectively to $0, 1, \infty, z$ for some z. Up to mirroring with the orientation-reversing reflection $z \mapsto \bar{z}$ we can suppose that $\Im z > 0$.
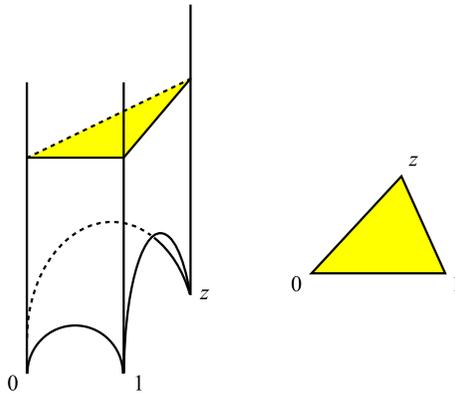
Figure 14.1. An ideal tetrahedron with three vertices in $0, 1, \infty$ in the half-space mode is determined by the position $z \in \mathbb{C} \cup \{\infty\}$ of the fourth vertex. A small horosphere centred at the ideal vertex intersects the tetrahedron in a Euclidean triangle uniquely determined up to similarities.
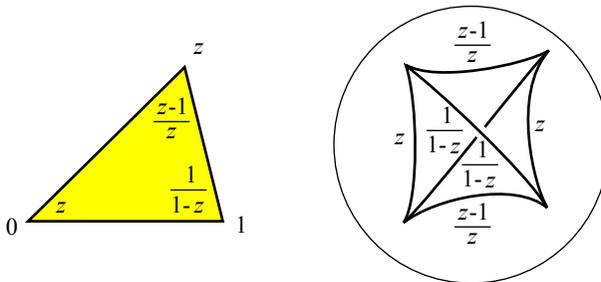
Figure 14.2. At each ideal vertex we have a Euclidean triangle defined up to similarities: each vertex of the triangle has a well-defined complex angle (left). We can assign the complex angles directly to the edges of the tetrahedron. The argument is the dihedral angle of the edge (right).

Remark 14.1.1. By definition the number z is the *cross-ratio* of the four complex numbers v_1, v_2, v_3, v_4.

A horosphere centred at the vertex $v_3 = \infty$ is a horizontal Euclidean plane that intersects the ideal tetrahedron in a Euclidean triangle as in Figure 14.1. The oriented similarity class of the triangle depends only on the vertex v_4, because a horosphere change results in a dilation: we can represent it as a triangle in $\mathbb{C} = \mathbb{R}^2$ with vertices at 0, 1, and z as in Figure 14.2-(left). The *complex angle* of a vertex of the triangle is the ratio of
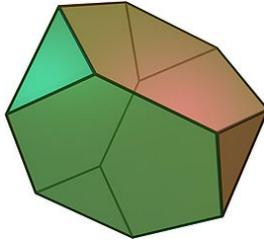
Figure 14.3. A truncated tetrahedron.

the two adjacent sides, taken with clockwise order and seen as complex
numbers. The three complex angles shown in Figure 14.2-(left) are:

$$z, \ \frac{1}{1-z}, \ \frac{z-1}{z}.$$

The argument is the usual angle, and the modulus is the ratio of the two
lengths of the adjacent sides.

Proposition 13.1.2 shows that some symmetries of an ideal tetrahe-
dron act on its vertices like the alternating group A_4 and hence transi-
tively. Therefore every vertex has the same triangular section as in Figure
14.2-(left), and all the sections can be encoded by assigning the complex
numbers directly to the edges of Δ as shown in Figure 14.2-(right). These
labels on the edges are called *moduli* and determine the ideal tetrahedron
up to orientation-preserving isometries of $\mathbb{H}^3$. The argument of a modulus
is the dihedral angle of the edge.

On a regular ideal tetrahedron the triangular sections are equilateral
and hence all edges have the same modulus $z = e^{\frac{\pi i}{3}}$.

14.1.2. Ideal triangulations. Let $\Delta_1, \ldots, \Delta_n$ be identical copies of
the standard oriented 3-simplex. As in the two-dimensional case, a *trian-
gulation* $\mathcal{T}$ is a partition of the $4n$ faces of the tetrahedra into $2n$ pairs, and
for each pair a simplicial isometry between the two faces. The triangulation
is *oriented* if the simplicial isometries are orientation-reversing. If we glue
the tetrahedra along the simplicial isometries we get a topological space
X, which is not necessarily a topological manifold. Let M be X minus the
vertices of the triangulation: we say that $\mathcal{T}$ is an *ideal triangulation* for M.

Proposition 14.1.2. *If $\mathcal{T}$ is oriented, then M is homeomorphic to the
interior of a compact oriented 3-manifold with boundary.*

Proof. To prove that M is a manifold we only need to check that a
point $x \in e$ in an edge e has a neighbourhood homeomorphic to an open
ball. A cycle of tetrahedra is attached to e, and since $\mathcal{T}$ is oriented we are
certain that a neighbourhood of x is a cone over a 2-sphere and not over
a projective plane.
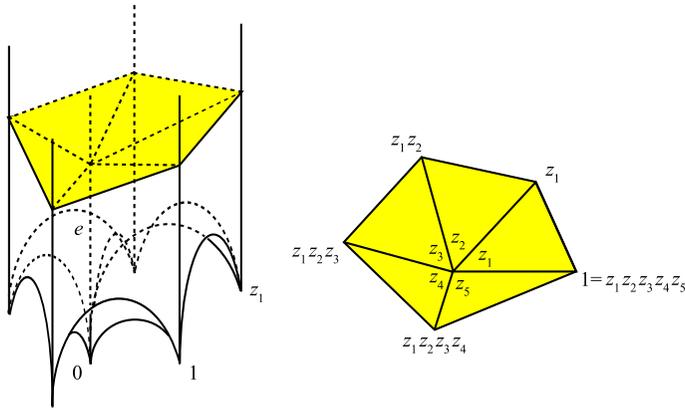
Figure 14.4. If we manage to glue all the tetrahedra incident to an edge e inside $\mathbb{H}^3$ as shown, the hyperbolic structure is defined also in e (left). Let $z_1, \ldots, z_h$ be the complex moduli assigned to the sides of the h incident tetrahedra (here $h = 5$). We can glue everything in $\mathbb{H}^3$ if and only if $z_1 \cdots z_h = 1$ and the arguments sum to 2π (right).

If we truncate the tetrahedra as in Figure 14.3 before gluing them, we get a compact manifold $N \subset M$ with boundary such that $M \setminus N \cong \partial N \times [0, 1)$. Therefore M is homeomorphic to $\text{int}(N)$. $\qquad\square$

We will always suppose that $\mathcal{T}$ is oriented and M is connected, so $M = \text{int}(N)$ for some compact N with boundary. Every ideal vertex v in $\mathcal{T}$ is locally a cone over a small triangulated closed surface $\Sigma \subset M$ obtained by truncating the tetrahedra incident to v. We call such a Σ the *link* of v.

14.1.3. Hyperbolic ideal triangulations. Let $\mathcal{T}$ be an oriented ideal triangulation with tetrahedra $\Delta_1, \ldots, \Delta_n$ of a 3-manifold M. We now substitute every Δ_i with an ideal hyperbolic tetrahedron and pair their faces with orientation-reversing isometries. As opposite to the two-dimensional case, the ideal hyperbolic tetrahedra are not unique (they depend on a complex modulus z_i) but the isometric pairing of their faces is uniquely determined in virtue of the following.

Proposition 14.1.3. *Given two ideal triangles Δ and Δ', every bijection between the ideal vertices of Δ and of Δ' is realised by a unique isometry.*

Proof. We see the ideal triangles in $\mathbb{H}^2$ and recall that for any two triples of points in $\partial \mathbb{H}^2$ there is a unique isometry sending pointwise the first triple to the second. Alternatively, we may use the barycentric decomposition shown in Figure 7.15-(left). $\qquad\square$

If we substitute each Δ_i with an ideal hyperbolic tetrahedron, we immediately get a well-defined hyperbolic structure on M minus the edges of $\mathcal{T}$. We now try to extend this hyperbolic structure to the edges: we can do this if we are able to glue all the h tetrahedra around each edge e inside $\mathbb{H}^3$ as in Figure 14.4. Let $z_1, \ldots, z_h$ be the complex moduli associated to the edges of the h tetrahedra incident to e. As shown in the figure, if $z_1 \cdots z_h = 1$ and the sum of their argument is 2π (and not some higher multiple of 2π) then all tetrahedra can be glued simultaneously inside $\mathbb{H}^3$ and the hyperbolic structure extends naturally to e. If this holds at every edge e of $\mathcal{T}$ then M inherits a hyperbolic structure and $\mathcal{T}$ is said to be a *hyperbolic* (or *geometric*) *ideal triangulation* for M.

14.1.4. Consistency equations. We want to parametrize the hyperbolic structures on M that may be constructed in this way from a fixed ideal triangulation $\mathcal{T}$. We pick an arbitrary edge for every tetrahedron Δ_i and assign to it a complex variable z_i with $\Im z_i > 0$, and the other edges of Δ_i are automatically labeled by one of the variables z_i, $\frac{z_i-1}{z_i}$, or $\frac{1}{1-z_i}$ as indicated in Figure 14.2 (recall that the tetrahedra are oriented). As we have seen, for every edge e in $\mathcal{T}$ we obtain an equation of type

$$w_1 \cdots w_h = 1$$

(to which we must add the condition that the sum of the arguments is 2π), where each w_j equals z_i, $\frac{z_i-1}{z_i}$, or $\frac{1}{1-z_i}$ for some i.

We have thus obtained a system of *consistency equations*, with a variable z_i for each tetrahedron and an equation $w_1 \cdots w_h = 1$ for each edge. A solution $z = (z_1, \ldots, z_n)$ to these equations produces a hyperbolic ideal triangulation and hence a hyperbolic structure on M. Recall that we assume $\Im z_i > 0 \quad \forall i$.

As in the two-dimensional case, the resulting hyperbolic structure is not necessarily complete, and to get a complete hyperbolic manifold we must add more equations.

14.1.5. Completeness equations. Our aim is to construct a complete finite-volume hyperbolic metric on M. By Corollary 4.2.18 if M has such a metric the link of every ideal vertex of $\mathcal{T}$ is a triangulated torus and identifies a cusp of M, so we will henceforth suppose that the links of all vertices are tori. In other words M is the interior of a compact 3-manifold N bounded by some $c > 0$ tori.

Let $z = (z_1, \ldots, z_n)$ be a solution to the consistency equations, providing a hyperbolic structure to M. Every boundary torus $T \subset \partial N$ is triangulated by $\mathcal{T}$: every triangle in T is the truncation triangle of some Δ_i and hence inherits the complex moduli of the three adjacent edges of Δ_i as in Figure 14.5-(left), thus it has a Euclidean structure well-defined up to similarities. We want to study the following problem: do these Euclidean structures on the triangles glue to form a Euclidean structure on
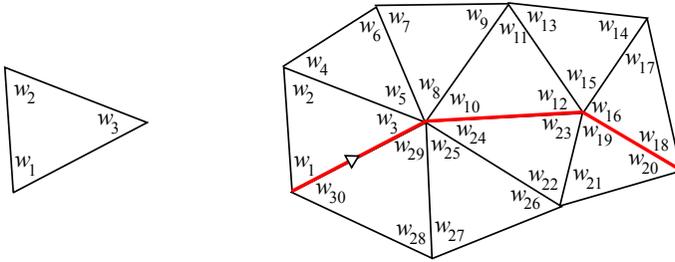
Figure 14.5. Every boundary torus $T \subset \partial N$ is triangulated, and each triangle has a Euclidean structure well-defined up to similarities and inherits three complex moduli w_1, w_2, w_3 at its vertices (left). At every vertex of the triangulation, the product of the adjacent moduli is 1, for instance here $w_{12}w_{15}w_{16}w_{19}w_{23} = 1$. The red path contributes to $\mu(\gamma)$ with the factor $-w_{30}w_{29}w_{25}w_{24}w_{23}w_{19}w_{20}$ (right).
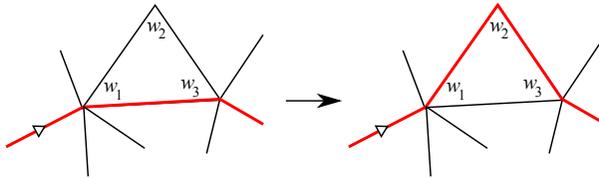
Figure 14.6. This move for γ does not affect $\mu(\gamma)$.

T? We will see that M is complete if and only if the answer is "yes" at each boundary torus T.

Pick $\gamma \in \pi_1(T) = H_1(T, \mathbb{Z}) \cong \mathbb{Z}^2$. We represent γ as a simplicial path in the triangulation of T and then define $\mu(\gamma) \in \mathbb{C}^*$ to be $(-1)^{|\gamma|}$ times the product of all the complex moduli that γ encounters at its right side, with $|\gamma|$ being the number of edges of γ, see Figure 14.5-(right).

Proposition 14.1.4. *The element $\mu(\gamma)$ is well-defined and $\mu \colon \pi_1(T) \to \mathbb{C}^*$ is a homomorphism.*

Proof. Two different paths for γ are related by moves as in Figure 14.6. This move does not affect $\mu(\gamma)$ since $w_1 w_2 w_3 = -1$ and the product of the moduli around a vertex is $+1$. The map μ is clearly a homomorphism. $\square$

Let $C(T) \subset M$ be a closed collar of the torus T in N, intersected with M. It is diffeomorphic to $T \times [0, +\infty)$.

Proposition 14.1.5. *The following facts are equivalent:*
  (1) *the homomorphism μ is trivial,*

(2) *there is a Euclidean structure on T that induces all the moduli,*
(3) *the manifold $C(T)$ is complete and contains a truncated cusp.*

Proof. The equivalence $(1) \Leftrightarrow (2)$ is a simple exercise. If (2) holds we may choose small horosections of all the hyperbolic tetrahedra incident to v that match to give a Euclidean torus $T_* \subset C(T)$. The non-compact part of $C(T)$ bounded by T_* is a truncated cusp with base T_* and is thus complete. On the other hand, if (1) does not hold, there is a path of horosections from a tetrahedron to itself which ends at a bigger height and we conclude that $C(T)$ is not complete as in the two-dimensional case, see the proof of Proposition 7.4.8. □

Corollary 14.1.6. *The hyperbolic manifold M is complete $\Longleftrightarrow$ μ is trivial for every torus $T \subset \partial N$.*

Fix for every boundary torus T two generators m, l for $H_1(T, \mathbb{Z})$. The homomorphism μ is trivial $\Longleftrightarrow$ the following two equations are satisfied:

$$\mu(m) = 1, \quad \mu(l) = 1.$$

Each equation is of some type $w_1 \cdots w_k = 1$. We get two equations for each of the c boundary tori and hence $2c$ equations in total, called the *completeness equations* for the triangulation $\mathcal{T}$. We summarise our discussion:

Proposition 14.1.7. *Let $\mathcal{T}$ be an ideal triangulation of $M = \mathrm{int}(N)$ with n tetrahedra and ∂N consisting of c tori. If a point $z = (z_1, \ldots, z_n)$ with $\Im z_i > 0$ satisfies the n consistency equations and the $2c$ completeness equations, then M admits a finite-volume complete hyperbolic metric.*

14.1.6. Examples. On a triangulation $\mathcal{T}$, the *valence* of an edge e is the number of tetrahedra incident to it, counted with multiplicity. The simplest kinds of solutions arise in the following construction.

Proposition 14.1.8. *Let $\mathcal{T}$ be a triangulation where all edges have valence six. The point $(e^{\frac{\pi i}{3}}, \ldots, e^{\frac{\pi i}{3}})$ is a solution of both the consistency and completeness equations and defines a hyperbolic structure where all the ideal tetrahedra are regular.*

Proof. Note that when $z = e^{\frac{\pi i}{3}}$ the tetrahedron is regular and we get

$$z = \frac{z-1}{z} = \frac{1}{1-z} = e^{\frac{\pi i}{3}}.$$

Therefore the moduli are $e^{\frac{\pi i}{3}}$ everywhere. The consistency equations are satisfied since $(e^{\frac{\pi i}{3}})^6 = 1$ and also the completeness equations are, because by assigning length 1 to each edge of the triangulation of a torus T we get a global Euclidean structure on T tessellated into isometric equilateral triangles. □
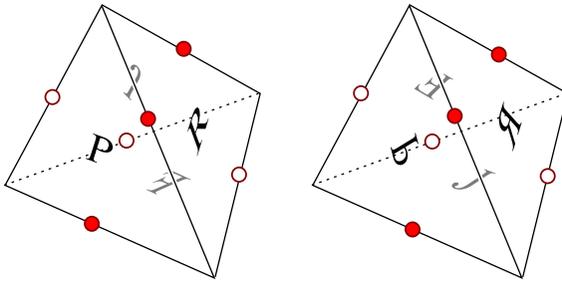
Figure 14.7. An oriented triangulation with 2 tetrahedra: faces with the same letter *F*, *J*, *P*, *R* are paired with a simplicial map that matches the letters. We get two edges with valence six (the red and white dots).
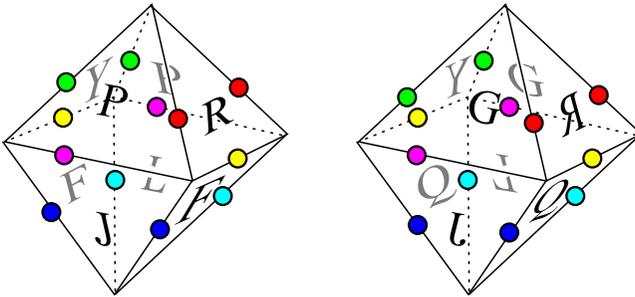
Figure 14.8. By pairing the faces of two ideal regular octahedra as shown we get a cusped complete hyperbolic three-manifold. Faces with the same letter are paired with the unique isometry that matches the letters. We get six edges with valence four (marked with coloured dots).

Figure 14.7 shows an ideal triangulation with two tetrahedra and two edges, each with valence six: this defines a cusped finite-volume complete hyperbolic three-manifold M that decomposes into two regular ideal tetrahedra.

Every finite covering $\tilde{M}$ of M is another example: the ideal triangulation $\mathcal{T}$ lifts to an ideal triangulation $\tilde{\mathcal{T}}$ where all the edges have valence six. Since the fundamental group $\pi_1(M)$ is residually finite (see Proposition 4.3.9) there are plenty of such coverings.

Exercise 14.1.9. Construct a triangulation $\mathcal{T}$ with one tetrahedron and one edge. The edge has valence six and hence the construction of Proposition 14.1.8 produces a cusped hyperbolic manifold. Note however

that the triangulation is not orientable and a cusp section is a Klein bottle! This non-orientable hyperbolic manifold is called the *Gieseking manifold*.

Similar kinds of "regular" examples may be constructed using some other platonic solids. The ideal regular octahedron is particularly interesting and useful because it is right-angled, see Section 3.2.4. If we pick finitely many oriented regular ideal octahedra $O_1, \ldots, O_h$ and pair orientation-reversingly their triangular faces in a way that every resulting edge has valence four, we construct a finite-volume cusped hyperbolic manifold M because $4 \times \frac{\pi}{2} = 2\pi$. An example with two octahedra is shown in Figure 14.8.

We can further decompose each regular ideal octahedron into four non-regular ideal tetrahedra (there are three ways to do this: you must choose a diagonal connecting two opposite vertices) to get an ideal triangulation $\mathcal{T}$ for M. The moduli of each such tetrahedron are

$$i, \ \frac{1}{1-i} = \frac{1+i}{2}, \ \frac{i-1}{i} = 1+i$$

and they satisfy the consistency and completeness equations in a less trivial way than before. The edges of the resulting triangulation have varying valences 4, 6, and 8.

14.1.7. SnapPea. It is of course hard to construct ideal triangulations by hand, except in some very symmetric and simple cases like the one just described. And it is even more difficult to solve the consistency and completeness equations.

There is a beautiful computer program that does all this for you! This is *SnapPea*, written by Jeff Weeks in the 1980s. Using SnapPea you can draw any link diagram L and the program immediately constructs an ideal triangulation for the complement of L in S^3. Then it uses the Newton method to find a numerical solution to the consistency and completeness equations: the solution is only numerical, but then one can use some *a posteriori* argument to confirm it rigorously.

For instance, if we draw the trefoil knot, SnapPea triangulates its complement and finds no solution to the equations: this is not surprising since its complement is Seifert and hence not hyperbolic, see Proposition 11.2.5.

If we draw the figure-eight knot in Figure 14.9-(left), SnapPea constructs precisely the ideal triangulation of Figure 14.7, and we have hence discovered that the complement of the figure-eight knot is hyperbolic and decomposes into two ideal regular tetrahedra.

We may draw links with many components, and discover for instance that the complement of the Borromean rings from Figure 14.9-(right) is hyperbolic and decomposes into two right-angled ideal regular octahedra. We can also Dehn-fill some boundary components to obtain more cusped manifolds that are not link complements in S^3.
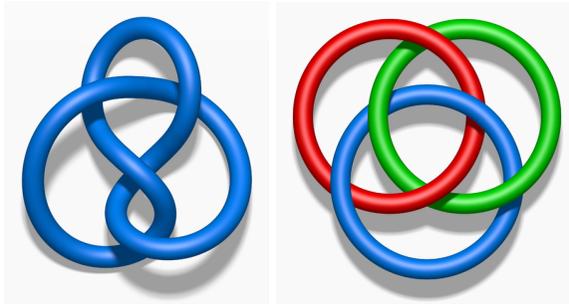
Figure 14.9. The figure-eight knot and the Borromean rings are *hyperbolic*, that is their complements admit complete hyperbolic metrics.

tetrahedra	1	2	3	4	5	6	7
1 cusp		2	9	52	223	913	3388
2 cusps				4	11	48	162
3 cusps						1	2

Table 14.1. The number of cusped orientable hyperbolic three-manifolds that can be triangulated with at most 7 ideal tetrahedra.

SnapPea can calculate numerically a wealth of geometrical invariants with some precision, including the volume and the first segment of the geodesic spectrum. It can manipulate manifolds (Dehn-filling, drilling along short simple closed geodesics, finite covers) and change triangulations. It provides beautiful pictures of the Dirichlet domain and of the cusp shapes.

Is SnapPea guaranteed to find a hyperbolic structure on M if there is one? No, it is not, although in practice it succeeds most of the time. The fact that M is hyperbolic does not guarantee that there is a solution to Thurston's equations on a given triangulation $\mathcal{T}$, and the existence of a solution does not guarantee that SnapPea will be able to find it.

We note that for most ideal triangulations $\mathcal{T}$ of a hyperbolic M there is no solution to Thurston's equations, and SnapPea is clever enough to modify $\mathcal{T}$ to increase the probability to find one. In all known examples, a cusped hyperbolic M admits at least *one* geometric ideal triangulation, but whether this holds true for all cusped hyperbolic 3-manifolds M is still an open question.

14.1.8. Cusped census. As with knots tabulations, topologists have used computers to list the cusped hyperbolic manifolds that can be triangulated with few ideal tetrahedra.

Name	Volume	Homology	Symmetry	SG	C
$M2_1$	2.0298832128	$\mathbb{Z}$	D_4	1.09	a
$M2_2$	2.0298832128	$\mathbb{Z} + \mathbb{Z}_5$	$\mathbb{Z}_2 + \mathbb{Z}_4$	0.86	a
$M3_1$	2.5689706009	$\mathbb{Z} + \mathbb{Z}_5$	D_4	0.65	c
$M3_2$	2.5689706009	$\mathbb{Z} + \mathbb{Z}_3$	D_4	0.65	c
$M3_3$	2.6667447834	$\mathbb{Z} + \mathbb{Z}_2$	D_4	0.63	c
$M3_4$	2.6667447834	$\mathbb{Z} + \mathbb{Z}_6$	D_4	0.63	c
$M3_5$	2.7818339124	$\mathbb{Z}$	$\mathbb{Z}_2$	0.51	c
$M3_6$	2.8281220883	$\mathbb{Z}$	D_4	0.56	c
$M3_7$	2.8281220883	$\mathbb{Z}$	$\mathbb{Z}_2$	0.58	c
$M3_8$	2.8281220883	$\mathbb{Z} + \mathbb{Z}_7$	D_4	0.56	c
$M3_9$	2.9441064867	$\mathbb{Z}$	$\mathbb{Z}_2$	0.43	c

Table 14.2. The 1-cusped orientable hyperbolic manifolds that can be triangulated with at most 3 ideal tetrahedra. The name Mi_j indicates that the manifold is the one with j-th smallest volume among those that can be triangulated with i tetrahedra. The column SG shows the length of the shortest closed geodesic (volume and SG values are truncated after few digits). The column C indicates whether the manifold is achiral (a) or chiral (c), that is if it admits an orientation-reversing isometry or not.

Name	Volume	Homology	Symmetry	SG	C
$M4_1^2$	3.6638623767	$\mathbb{Z} + \mathbb{Z}$	D_8	1.06	c
$M4_2^2$	3.6638623767	$\mathbb{Z} + \mathbb{Z}$	D_8	0.96	c
$M4_3^2$	4.0597664256	$\mathbb{Z} + \mathbb{Z}$	D_{12}	0.86	c
$M4_4^2$	4.0597664256	$\mathbb{Z} + \mathbb{Z}$	D_8	0.86	a

Table 14.3. The 2-cusped orientable hyperbolic manifolds that can be triangulated with 4 ideal tetrahedra. The name Mi_j^k indicates the k-cusped manifold with j-th smallest volume among those that are triangulated with i tetrahedra.

The number of cusped hyperbolic orientable manifolds that can be ideally triangulated with n tetrahedra (and not less than n) is written in Table 14.1 for all $n \leqslant 7$. The 1-cusped manifolds with $n = 2, 3$ are listed with more detail in Table 14.2 and the 2-cusped ones with $n = 4$ are in Table 14.3. These tables were produced by Callahan − Hildebrand − Weeks [**10**] in 1999 via a computer enumeration.

The figure-eight knot complement is $M2_1$, while $M2_2$ is another hyperbolic manifold that decomposes into two regular ideal tetrahedra: it is not a knot complement in S^3 since its homology is not $\mathbb{Z}$, but it is
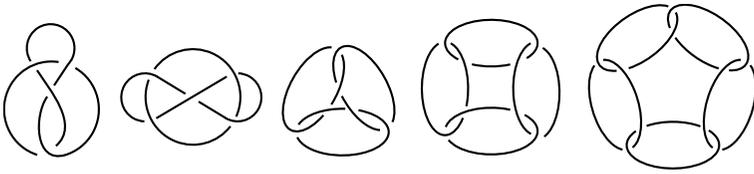
Figure 14.10. A remarkable sequence of hyperbolic links. These are the *figure eight knot*, the *Whitehead link*, and some particular *chain links* with 3,4,5 components.

the complement of a knot K in the lens space $L(5, 1)$, called the *figure-eight knot sibling*. This knot is obtained by performing a (-5)-surgery on one component of the *Whitehead link* shown in Figure 14.10: since this component is trivial, the surgered manifold is $L(5, 1)$ and the other unsurgered component is our knot K. Both manifolds $M2_1$ and $M2_2$ have volume $2.0298832128\ldots$ since the volume of the ideal regular tetrahedron is $3\Lambda\left(\frac{\pi}{3}\right) = 1.0149416064\ldots$

The manifold $M4_1^2$ is the Whitehead link complement, that can be obtained by pairing the faces of a single ideal regular octahedron O. Indeed the volume of O is $8\Lambda\left(\frac{\pi}{4}\right) = 3.6638623767\ldots$

After the figure eight knot and the Whitehead link, a notable sequence of hyperbolic links in S^3 with increasing number of components is shown in Figure 14.10. Their complements are conjectured to be the smallest hyperbolic manifolds with $i = 1, \ldots, 5$ cusps (see the references at the end of the chapter).

14.1.9. Hyperbolic knots. How many hyperbolic knots are there in the three-sphere? There are in fact infinitely many hyperbolic knots, and infinitely many non-hyperbolic knots. Geometrisation translates hyperbolicity into an appealing topological condition:

Theorem 14.1.10. *A knot $K \subset S^3$ is either a torus knot, a satellite knot, or a hyperbolic knot.*

Proof using geometrisation. Proposition 11.2.11 says that K is either a torus knot, a satellite knot, or has simple complement M. In the latter case M is hyperbolic by Corollary 12.9.6.   □

Note that the three cases are mutually exclusive. The number of prime knots with $c \leqslant 14$ crossings in each of the three classes is shown in Table 14.4, taken from Hoste − Thistlethwaite − Weeks [**29**]: the table shows a strong predominance of hyperbolic knots, but it is unknown whether it persists also when c tends to infinity.

c	3	4	5	6	7	8	9	10	11	12	13	14
torus	1	0	1	0	1	1	1	1	1	1	0	1	1
satellite	0	0	0	0	0	0	0	0	0	0	2	2
hyperbolic	0	1	1	3	6	20	48	164	551	2176	9985	46969

Table 14.4. The number of torus, satellite, and hyperbolic
prime knots with c crossings, for all $c \leqslant 14$.

14.2. Closed hyperbolic three-manifolds

We have described in Section 14.1 a method to construct cusped
finite-volume hyperbolic three-manifolds. How can we now build some
closed hyperbolic three-manifolds?

Following the same path, we could try to parametrize *compact* hyper-
bolic tetrahedra via some variables, and then encode the isometric gluings
of their faces via some equations. However, it is really hard to parametrize
compact tetrahedra and their gluings, and nobody has ever constructed
any closed hyperbolic three-manifold in this way!

The usual procedure for building closed hyperbolic three-manifolds
consists of using (again) ideal tetrahedra: we start with a cusped manifold,
and then we slightly modify the completeness equations to allow some
appropriate Dehn-filling of some (or all) cusps.

14.2.1. Dehn filling parameters. We now consider an orientable 3-
manifold $M = \mathrm{int}(N)$ where N is compact and $\partial N = T_1 \sqcup \ldots \sqcup T_c$ consists
of tori. For simplicity we will always assume that each T_i is incompressible,
so that $\pi_1(T_i) = \mathbb{Z} \times \mathbb{Z}$ injects in $\pi_1(N) = \pi_1(M)$ for all i.

We fix once for all two generators m_i, l_i of $\pi_1(T_i)$ for each $i = 1, \ldots, c$.
A *Dehn filling parameter* $s = (s_1, \ldots, s_c)$ is a sequence where each s_i is
either a pair (p, q) of coprime integers or the symbol ∞. It is useful to
think of s_i as a point in the two-sphere $S^2 = \mathbb{R}^2 \cup \{\infty\}$.

A Dehn filling parameter s determines a Dehn filling of N as follows:
for every i, if $s_i = (p, q)$ we fill T_i by killing the slope $pm_i + ql_i$, while
if $s_i = \infty$ we do nothing. The result is a new compact manifold, whose
interior we denote by M^{fill}, that may be closed (if there are no ∞ in s) or
may be bounded by some tori.

We fix a Dehn filling parameter s, producing a filled manifold M^{fill}.

14.2.2. The modified equations. Let $\mathcal{T}$ be an oriented ideal triangu-
lation for M. We now describe some equations similar to that of Section
14.1, whose solutions now identify hyperbolic structures on M^{fill} instead of
M. The equations of course will depend on the Dehn filling parameter s,
since the manifold M^{fill} does.

The triangulation $\mathcal{T}$ consists of some n tetrahedra $\Delta_1, \ldots, \Delta_n$; we choose an edge in each Δ_i and we assign it the variable $z_i \in \mathbb{C}$ with $\Im z_i > 0$ as we did in Section 14.1.

The edges of all tetrahedra are coloured with moduli z_i, $\frac{1}{1-z_i}$, $\frac{z_i - 1}{z_i}$ that lie in the upper half-plane and as such may be written in polar coordinates as $\rho e^{i\theta}$ with $\theta \in (0, \pi)$. It is convenient to consider these variables as elements of the multiplicative group

$$\widetilde{\mathbb{C}^*} = \left\{ \rho e^{i\theta} \mid \rho \in \mathbb{R}_{>0}, \ \theta \in \mathbb{R} \right\}$$

that covers $\mathbb{C}^*$ and is isomorphic to $\mathbb{C}$ via the map $\exp \colon \mathbb{C} \to \widetilde{\mathbb{C}^*}$.

Every edge of $\mathcal{T}$ furnishes a *consistency equation* in $\widetilde{\mathbb{C}^*}$ as described in Section 14.1, of the form

$$w_1 \cdots w_h = e^{2\pi i}.$$

Note that $e^{2\pi i} \neq 1$ in $\widetilde{\mathbb{C}^*}$, so by interpreting the consistency equations in $\widetilde{\mathbb{C}^*}$ we have also incorporated the request that all angles sum to 2π.

Let $z = (z_1, \ldots, z_n)$ be a solution to the consistency equations. We have defined in Proposition 14.1.4 a homomorphism $\mu \colon \pi_1(T_i) \to \mathbb{C}^*$ for every boundary torus T_i, and we now lift it to $\widetilde{\mathbb{C}^*}$.

Definition 14.2.1. Represent each non-trivial $\gamma \in \pi_1(T_i)$ as a simplicial path that lifts to an embedded path in the universal cover of T_i. Define $\tilde{\mu}(\gamma)$ to be $e^{-|\gamma|\pi i}$ times the product of all the moduli that γ encounters at its right side, considered now as elements in the group $\widetilde{\mathbb{C}^*}$. If γ is trivial, set $\tilde{\mu}(\gamma) = 1$.

Exercise 14.2.2. The element $\tilde{\mu}(\gamma) \in \widetilde{\mathbb{C}^*}$ is well-defined and

$$\tilde{\mu} \colon \pi_1(T_i) \to \widetilde{\mathbb{C}^*}$$

is a homomorphism.

We now define some new *completeness equations* at T_i for each $i = 1, \ldots, c$ that depend on the Dehn filling parameter s_i. There are two cases to consider for each $i = 1, \ldots, c$. If $s_i = \infty$, the completeness equations relative to T_i are the two equations

$$\mu(m_i) = \mu(l_i) = 1$$

already considered in Section 14.1.5. (If we wish, we can substitute them with the equations $\tilde{\mu}(m_i) = \tilde{\mu}(l_i) = 1$. Both choices will work.)

If $s_i = (p, q)$ we define one new completeness equation

$$\tilde{\mu}(m_i)^p \cdot \tilde{\mu}(l_i)^q = e^{2\pi i}.$$

The total number of completeness equations therefore varies from c to $2c$, depending on the number of cusps that are left unfilled. The main goal of this section is to prove the following.

Theorem 14.2.3. *A solution $z = (z_1, \ldots, z_n)$ to the consistency and completeness equations determines a hyperbolic structure on M whose completion $\overline{M}$ is a complete hyperbolic manifold diffeomorphic to M^{fill}.*

If $s \neq (\infty, \ldots, \infty)$ the solution z determines an incomplete hyperbolic metric on M, and the miracle here is that the completion $\overline{M}$ is another hyperbolic manifold!

We note that both the consistency and completeness equations are of type $w_1 \cdots w_h = 1$ or $e^{2\pi i}$. The rest of this section is mainly devoted to the proof of Theorem 14.2.3. We start with a short discussion that introduces (truncated) solid tori with cone angles: these objects will be crucial in the proof of the theorem.

14.2.3. The infinite branched covering. Let $l \subset \mathbb{H}^3$ be any line and recall the universal cover

$$\pi \colon X \longrightarrow \mathbb{H}^3 \setminus l$$

already considered in Section 3.5.4 when we introduced manifolds with cone singularities. The manifold X is incomplete and its completion $\overline{X}$ is obtained by adding a copy $\tilde{l}$ of l.

We use the half-space model and represent l as the vertical axis, so that $\mathbb{H}^3 \setminus l = \mathbb{C}^* \times \mathbb{R}_{>0}$ and we can write

$$X = \widetilde{\mathbb{C}^*} \times \mathbb{R}_{>0}.$$

Note that $\widetilde{\mathbb{C}^*}$ acts on X via isometries: the element $w \in \widetilde{\mathbb{C}^*}$ acts as

$$(z, t) \longmapsto (wz, |w|t).$$

When $w = e^{\alpha i}$ this is the rotation of angle α that was used to define hyperbolic cone manifolds. When $|w| \neq 1$ the map projects to a hyperbolic isometry of $\mathbb{H}^3$ with axis l and translation distance $\log|w|$.

14.2.4. Tubes with cone angles. A non-trivial discrete group $\Gamma < \widetilde{\mathbb{C}^*} \cong \mathbb{C}$ is either isomorphic to $\mathbb{Z}$ or to $\mathbb{Z} \times \mathbb{Z}$; it acts freely on X and (not necessarily freely!) on the line $\tilde{l}$. Concerning $\mathbb{Z} \times \mathbb{Z}$, there are two cases to consider.

If $\Gamma = \mathbb{Z} \times \mathbb{Z}$ contains a non-trivial rotation, it is generated by some maps

$$(z, t) \mapsto (e^{\alpha i} z, t), \quad (z, t) \mapsto e^{\lambda}(e^{\beta i} z, t)$$

with $\lambda \neq 0$, and it acts on $\tilde{l}$ as translations whose step is an integer multiple of λ. The quotient $\overline{X}/_\Gamma$ is naturally a complete hyperbolic manifold with cone angles: it is an open solid torus with singular locus a closed geodesic of length $|\lambda|$ and with cone angle α. We call it a *tube with cone angle* α. When $\alpha = 2\pi$ we get an ordinary tube as in Section 4.1.2, with no singular points.

If $\Gamma = \mathbb{Z} \times \mathbb{Z}$ does not contain a rotation, it acts on $\tilde{l}$ as an indiscrete group of translations and hence $\overline{X}/_\Gamma$ is not a hyperbolic manifold with cone angles (it is not even Hausdorff). We also note that $\overline{X}/_\Gamma$ is not the

completion of the hyperbolic manifold $X/_\Gamma$ in that case, but it only maps surjectively onto it: the completion adds a single point to $X/_\Gamma$.

Tubes have natural truncations. For every $R > 0$ the R-neighbourhood $N_R(\tilde{l}) \subset \overline{X}$ of the singular line $\tilde{l}$ projects onto the R-neighbourhood $N_R(l)$ of l, which is a Euclidean cone with axis l. The group Γ preserves $N_R(\tilde{l})$ and in the first case its quotient is a truncated tube with cone angles.

14.2.5. Incomplete solutions. We now go back to our triangulation $\mathcal{T}$, and we let $z = (z_1, \ldots, z_n)$ be a solution to the consistency equations for $\mathcal{T}$. The solution z furnishes a hyperbolic metric on M, which might not be complete: we want to understand the metric completion $\overline{M}$ of M. Recall that M is diffeomorphic to the interior of a compact N whose boundary consists of c tori.

We pick a boundary torus $T \subset \partial N$ and define a *collar* $C(T) \subset M$ of T to be the intersection of a closed collar of T in N with M. We know from Proposition 14.1.5 that when $\mu\colon \pi_1(T) \to \mathbb{C}^*$ is trivial there is a collar $C(T)$ isometric to a truncated cusp, that is the truncated quotient of $\mathbb{H}^3$ by a discrete $\mathbb{Z} \times \mathbb{Z}$ of parabolic elements.

When μ is not trivial, we now show that a similar (but different) configuration arises: there is a collar $C(T)$ isometric to a truncated quotient of the hyperbolic manifold X considered above by a discrete $\mathbb{Z} \times \mathbb{Z}$. The crucial difference is that $C(T)$ now is incomplete, because X is. To prove this fact we use developing maps and holonomies.

The hyperbolic ideal triangulation $\mathcal{T}$ of M determined by z lifts to a hyperbolic ideal triangulation $\tilde{\mathcal{T}}$ for the universal cover $\tilde{M} \to M$ with infinitely many tetrahedra. The triangulated torus T is the link of a vertex v of $\mathcal{T}$. We fix a lift $\tilde{v}$ of v in $\tilde{\mathcal{T}}$, and the link of $\tilde{v}$ is a triangulated surface $\tilde{T}$ that covers T. Since T is incompressible, $\pi_1(T)$ injects in $\pi_1(M)$ and hence $\tilde{T}$ is a plane.

Recall from Section 3.5.3 that there is a developing map $D\colon \tilde{M} \to \mathbb{H}^3$ and a holonomy $\rho\colon \pi_1(M) \to \mathrm{Isom}^+(\mathbb{H}^3)$, and both are determined once we define D on an ideal tetrahedron in $\tilde{\mathcal{T}}$. We choose for our convenience one ideal tetrahedron incident to $\tilde{v}$ and we map it to the half-space model H^3 with an isometry D that sends $\tilde{v}$ to ∞.

The developing map D sends all tetrahedra of $\tilde{\mathcal{T}}$ incident to $\tilde{v}$ to vertical ideal tetrahedra in H^3 with one vertex at ∞. If $\gamma \in \pi_1(T)$, then $\rho(\gamma)$ permutes these tetrahedra and fixes ∞. Therefore the holonomy ρ sends $\pi_1(T) = \mathbb{Z} \times \mathbb{Z}$ to a group of commuting elements in $\mathbb{PSL}_2(\mathbb{C})$ fixing ∞. Every such element may be written as $z \mapsto az + b$, and there are two possibilities:

- the group consists of translations $z \mapsto z + b$,
- the group fixes a point $p \in \mathbb{C}$.

In the second case we may suppose $p = 0$ up to translating everything, so the maps are all of type $z \mapsto az$. Not surprisingly, the holonomy ρ is tightly connected with the homomorphism $\mu\colon \pi_1(T) \to \mathbb{C}^*$.

Exercise 14.2.4. The derivative $\rho(\gamma)'$, that is the coefficient 1 or a in the above examples, equals $\mu(\gamma)$.

We now suppose that μ is non-trivial and hence the first possibility is excluded: the group $\rho(\pi_1(T))$ consists of maps $z \mapsto az$ and we can identify ρ with μ. The map D induces a developing map $D \colon \tilde{T} \to \mathbb{C}$ with holonomy μ.

Proposition 14.2.5. *The image $D(\tilde{T})$ misses the origin.*

Proof. The vector field $v(z) = z$ on $\mathbb{C}$ is $\mu(\pi_1(T))$-invariant and hence pulls-back via D to a vector field on T. If $0 \in D(\tilde{T})$ then this vector field on T has some zeroes, all with index 1: a contradiction since $\chi(T) = 0$.   □

Let $\mathrm{St}_{\tilde{v}} \subset \tilde{M}$ be the open star of $\tilde{v}$, that is the union of all ideal tetrahedra incident to $\tilde{v}$, with the faces not incident to $\tilde{v}$ removed: it is homeomorphic to $\tilde{T} \times (0, +\infty)$ and hence simply connected. The proposition implies that the restriction $D|_{\mathrm{St}_{\tilde{v}}}$ of the developing map misses the entire vertical line $l \subset \mathbb{H}^3$ above the origin and hence lifts to a map

$$\tilde{D} \colon \mathrm{St}_{\tilde{v}} \longrightarrow X.$$

Likewise the holonomy ρ at $\pi_1(T)$ lifts to a holonomy

$$\tilde{\rho} \colon \pi_1(T) \longrightarrow \widetilde{\mathbb{C}^*} < \mathsf{Isom}^+(X)$$

which is of course related to the homomorphism $\tilde{\mu} \colon \pi_1(T) \to \widetilde{\mathbb{C}^*}$.

Exercise 14.2.6. We have $\tilde{\rho} = \tilde{\mu}$.

After this long discussion, we can finally discover how a collar $C(T)$ of T looks like.

Proposition 14.2.7. *If μ is non-trivial then $\tilde{\mu}$ is injective with discrete image. There is a collar $C(T)$ that is isometric to a truncation of $X/_{\mathsf{Im}\,\tilde{\mu}}$.*

Proof. We have defined a lifted developing map $\tilde{D} \colon \mathrm{St}_{\tilde{v}} \to X$ with holonomy $\tilde{\rho} = \tilde{\mu}$.

Let $\Delta_{i_1}, \ldots, \Delta_{i_h}$ be the tetrahedra incident to v (with multiplicities), and let $\tilde{\Delta}_{i_1}, \ldots, \tilde{\Delta}_{i_h}$ be any lifts in $\tilde{\mathcal{T}}$ incident to $\tilde{v}$. We fix a sufficiently small $R_0 > 0$ such that the cone $N_{R_0}(\tilde{l}) \subset \overline{X}$ does not intersect the lower faces of the tetrahedra $\tilde{D}(\tilde{\Delta}_{i_1}), \ldots, \tilde{D}(\tilde{\Delta}_{i_h})$, see Figure 14.11-(left). Since $N_{R_0}(\tilde{l})$ is $\rho(\pi_1(T))$-invariant, it does not intersect the lower faces of any ideal tetrahedron in the image of $\tilde{D}$, and hence it intersects every ideal tetrahedron in a curved triangle as in Figure 14.11-(right).

We define $\widetilde{C(T)} \subset \tilde{M}$ as the preimage of $N_{R_0}(\tilde{l})$ along $\tilde{D}$. By construction it is a $\pi_1(T)$-invariant submanifold that projects to a submanifold $C(T) \subset M$. This submanifold intersects every tetrahedron incident to v
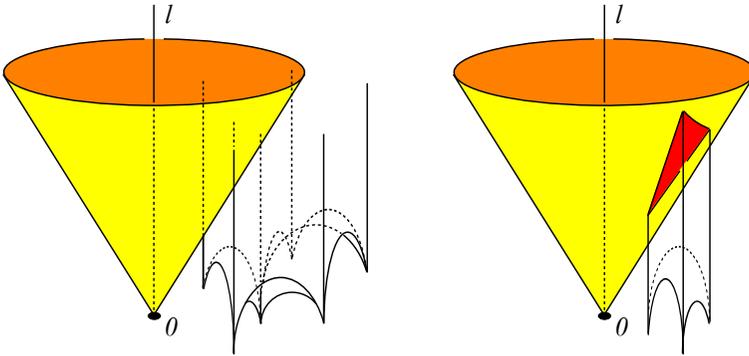
Figure 14.11. For sufficiently small $R > 0$ the cone neighbourhood $N_R(\tilde{l})$ of $\tilde{l}$ does not intersect the lower faces of the developed images of $\tilde{\Delta}_{i_1}, \ldots, \tilde{\Delta}_{i_h}$ (left). These tetrahedra intersect $\partial N_R(\tilde{l})$ into triangles that glue up to determine a torus T_R in $C(T)$ (right).

into a cone neighbourhood of v and is hence a collar for T. We want to prove that the map

$$\tilde{D} \colon \widetilde{C(T)} \longrightarrow N_{R_0}(\tilde{l}) \setminus \tilde{l}$$

is an isometry. Being a local isometry, it suffices to prove that it is injective and surjective. We prove this using the natural foliation of $N_{R_0}(\tilde{l}) \setminus \tilde{l}$ into the sheets $\partial N_R(\tilde{l})$ with $R \leqslant R_0$.

For every $R \leqslant R_0$ the map $\tilde{D}$ restricts to a local isometry of surfaces

$$\tilde{D}_R \colon \tilde{T}_R \to \partial N_R(\tilde{l})$$

where $\tilde{T}_R = \tilde{D}^{-1}(\partial N_R(\tilde{l}))$ is $\pi_1(T)$-invariant and covers a torus $T_R \subset M$ parallel to T. Since T_R is compact, the cover $\tilde{T}_R$ is complete. Therefore $\tilde{D}_R$ is a covering by Proposition 1.2.19 and is hence an isometry since $\partial N_R(\tilde{l})$ is simply-connected. In particular it is a bijection for all $R < R_0$, and hence $\tilde{D}$ is a bijection.

Since $\tilde{D}$ is an isometry, its holonomy $\tilde{\rho}$ is discrete and injective and

$$C(T) \cong \left(N_{R_0}(\tilde{l}) \setminus \tilde{l} \right)/_{\tilde{\rho}(\pi_1(T))}.$$

The proof is complete. $\square$

14.2.6. The completion. We now have an isometric model for the collar $C(T)$ of T and we study its completion. This analysis will lead to a proof of Theorem 14.2.3.

We suppose that μ is non-trivial and hence $C(T)$ is not complete. By Proposition 14.2.7, the completion $\overline{C(T)}$ of $C(T)$ depends on the $\mathbb{Z} \times \mathbb{Z}$ group $\text{Im}\,\tilde{\mu} < \widehat{\mathbb{C}^*}$ acting on X: see Section 14.2.4 where we defined in particular the (truncated) tubes with cone angles.

Proposition 14.2.8. *If* $\text{Im}\,\tilde{\mu}$ *contains a non-trivial rotation then* $\overline{C(T)}$ *is a truncated tube with some cone angle; otherwise it is a one-point compactification of* $C(T)$.

In the first case, there are generators γ, η *for* $\pi_1(T)$ *such that*

$$\tilde{\mu}(\gamma) = e^{\alpha i}, \qquad \tilde{\mu}(\eta) = e^{\lambda + \beta i}$$

and the core geodesic of $\overline{C(T)}$ *has cone angle* α *and length* $|\lambda|$. *The curve* γ *is a meridian of the truncated tube.*

Proof. We know that $C(T)$ is isometric to a truncation of $X/_{\text{Im}\,\tilde{\mu}}$ and we apply the discussion of Section 14.2.4. □

We now look at M globally: every boundary torus $T_i \subset \partial N$ has its own homomorphisms μ and $\tilde{\mu}$.

Corollary 14.2.9. *Suppose that at every boundary torus* $T_i \subset \partial N$ *one of the following holds:*

(1) μ *is trivial, or*
(2) *there is a primitive* $\gamma_i \in \pi_1(T_i)$ *such that* $\tilde{\mu}(\gamma_i) = e^{\alpha_i i}$.

The completion $\overline{M}$ *is a complete hyperbolic cone manifold obtained by Dehn filling the slopes* γ_i. *The core of the i-th Dehn filling is a closed geodesic with cone angle* α_i.

Theorem 14.2.3 now follows immediately.

Proof of 14.2.3. The equation $\tilde{\mu}(m_i)^p \tilde{\mu}(l_i)^q = e^{2\pi i}$ implies that $\tilde{\mu}(pm_i + ql_i) = e^{2\pi i}$ and therefore the completion is a hyperbolic manifold (with no cone angles since $\alpha = 2\pi$) diffeomorphic to M^{fill}. □

14.2.7. Generalised Dehn filling invariants. We have proved Theorem 14.2.3 and we now make some comments that will be useful in the next chapter. Let z be a solution of the consistency equations and consider a boundary torus $T_i \subset \partial N$ with fixed generators m_i, l_i for $\pi_1(T)$.

Proposition 14.2.10. *If* μ *is non-trivial, there is a unique* $(p, q) \in \mathbb{R}^2$ *such that* $\tilde{\mu}(m_i)^p \tilde{\mu}(l_i)^q = e^{2\pi i}$.

Proof. We know from Proposition 14.2.7 that $\tilde{\mu}$ is injective and has discrete image. Therefore $\tilde{\mu}(m_i)$ and $\tilde{\mu}(l_i)$ form a basis of $\widetilde{\mathbb{C}^*}$ considered as a $\mathbb{R}$-vector space. □

We now define the *generalised Dehn filling invariant* $(p, q) \in S^2 = \mathbb{R}^2 \cup \{\infty\}$ of the solution z at the torus T_i to be:

- $(p, q) = \infty$ if μ is trivial,
- the $(p, q) \in \mathbb{R}^2$ from Proposition 14.2.10 if μ is non-trivial.

If $(p, q) = \infty$ there is a complete collar $C(T_i)$ of T_i that is a truncated cusp; if $(p, q) \in \mathbb{R}^2$ then Proposition 14.2.7 provides a nice incomplete collar $C(T_i)$ and Proposition 14.2.8 tells us everything about its completion $\overline{C(T_i)}$. Namely, the following holds:

- if $\frac{p}{q} \in \mathbb{Q} \cup \{\infty\}$ then $(p, q) = k(r, s)$ for a unique real number $k > 0$ and coprime integers (r, s), and $\overline{C(T_i)}$ is a solid torus with meridian $rm + sl$, isometric to a truncated tube with cone angle $\frac{2\pi}{k}$;
- if $\frac{p}{q} \notin \mathbb{Q} \cup \{\infty\}$ then $\overline{C(T_i)}$ is the much less interesting one-point compactification.

We deduce the following:

- if (p, q) are coprime integers then $\overline{C(T_i)}$ is a standard truncated tube; if this holds at all boundary tori T_i then $\overline{M}$ is a hyperbolic manifold;
- if (p, q) are integers then k is a natural number and the cone angle divides 2π; if this holds at all boundary tori T_i then $\overline{M}$ may be interpreted as an orbifold thanks to Proposition 3.6.18;
- if $\frac{p}{q} \in \mathbb{Q} \cup \{\infty\}$ at all boundary tori T_i then $\overline{M}$ is a hyperbolic cone manifold.

Corollary 14.2.11. *If the generalised Dehn filling invariants (p, q) at each T_i are either ∞ or coprime integers, the completion $\overline{M}$ is a complete hyperbolic manifold obtained by (p, q)-Dehn filling the cusps of the second type.*

We end this discussion by calculating (p, q) explicitly. We define

$$u_i = \log \tilde{\mu}(m_i), \quad v_i = \log \tilde{\mu}(l_i)$$

and find the following.

Proposition 14.2.12. *If μ is non trivial we have*

$$p = -2\pi \frac{\Re v_i}{\Im(\bar{u}_i v_i)}, \qquad q = 2\pi \frac{\Re u_i}{\Im(\bar{u}_i v_i)}.$$

Proof. The pair (p, q) is the unique solution of

$$(20) \qquad\qquad p u_i + q v_i = 2\pi i$$

when $u_i, v_i \neq 0$, or is ∞ otherwise. Note that $u_i = 0 \Leftrightarrow v_i = 0$. The relation

$$i\Im(\bar{u}_i v) = -(\Re v_i) u_i + (\Re u_i) v_i$$

implies easily that the pair (p, q) stated above is a solution to (20). $\qquad\square$

14.2.8. Closed census. The Lickorish-Wallace Theorem 11.3.15 says that every closed orientable three-manifold M is the result of a Dehn surgery along some link $L \subset S^3$, and as such it can be easily presented to SnapPea. The program tries to solve numerically the consistency and completeness equations for M, based on some ideal triangulation for the complement of L.

If it succeeds to find a solution z, the closed manifold M is hyperbolic and SnapPea calculates numerically various geometric invariants of M: the volume (which is just the sum of the volumes of the ideal tetrahedra),

Name	Volume	Homology	Symmetry	SG	C
Vol1	0.94270736	$\mathbb{Z}_5 + \mathbb{Z}_5$	D_{12}	0.5846	c
Vol2	0.98136883	$\mathbb{Z}_5$	D_4	0.5780	c
Vol3	1.01494161	$\mathbb{Z}_3 + \mathbb{Z}_6$	S_{16}	0.8314	a
Vol4	1.26370924	$\mathbb{Z}_5 + \mathbb{Z}_5$	D_8	0.5750	c
Vol5	1.28448530	$\mathbb{Z}_6$	D_4	0.4803	c
Vol6	1.39850888	$\{e\}$	D_4	0.3661	c
Vol7	1.41406104	$\mathbb{Z}_6$	D_4	0.7941	c
Vol8	1.41406104	$\mathbb{Z}_{10}$	D_4	0.3648	c
Vol9	1.42361190	$\mathbb{Z}_{35}$	D_4	0.3523	c
Vol10	1.44069901	$\mathbb{Z}_3$	D_4	0.3615	c

Table 14.5. The ten closed hyperbolic three-manifolds with smallest volume known. Here S_{16} is the semidihedral group of order 16 with presentation $\langle x, y \mid x^8 = y^2 = 1, y^{-1}xy = x^3 \rangle$, SG indicates the length of the shortest geodesics (volume and SG values are truncated after few digits), and C indicates whether the manifold is achiral (a) or chiral (c), that is whether it admits an orientation-reversing isometry or not.

the length of the core geodesics of the filling solid tori (using Proposition 14.2.8), a segment of the geodesic spectrum, a Dirichlet domain, etc.

Various closed manifolds M have been tested and listed, and the ten closed hyperbolic three-manifolds of smallest volume known today are in Table 14.5, taken from Hodgson − Weeks [28].

14.2.9. References. The material contained in this chapter originated from Thurston's notes [56] and Neumann − Zagier [42], see also Benedetti − Petronio [4]. The program SnapPea is freely available and can now be used via a Python interface [15]. We have used the computer censuses of Callahan − Hildebrand − Weeks [10], Hoste − Thistlethwaite − Weeks [29], and Hodgson − Weeks [28].

It is conjectured in [1] that each link in Figure 14.10 with $c = 1, \ldots, 5$ cusps has minimum volume among orientable hyperbolic manifolds with c cusps. This has been proved for $c = 1$ by Cao and Meyerhoff [11], for $c = 2$ by Agol [1], and for $c = 4$ by Yoshida [61]. The manifold Vol1 from Table 14.5 is called the *Fomenko − Matveev − Weeks manifold* and has indeed smallest volume among all complete hyperbolic three-manifolds by Gabai − R. Meyerhoff − P. Milley [22].

Hyperbolic Dehn filling

We have completely classified in Chapter 10 the closed three-manifolds that belong to the six Seifert geometries, and now we long for a similar catalogue that displays all the closed hyperbolic three-manifolds that exist in nature. Is there something like a "name" to assign to each closed hyperbolic three-manifold, together with some reasonable tables that list all possible names?

There is not yet one such thing, and we do not know if there will ever be one: we still do not understand hyperbolic three-manifolds globally. The main difficulty is that there are really many hyperbolic three-manifolds around, so many that topologists often say informally that "most three-manifolds are hyperbolic."

This folk sentence is supported by a fundamental theorem that we prove in this chapter, the *Hyperbolic Dehn filling Theorem*, which says roughly that by Dehn-filling generically a cusped hyperbolic three-manifold we still get a hyperbolic manifold.

15.1. Introduction

It sometimes happens in three-dimensional topology, that a topological or geometric property of a manifold is preserved under Dehn fillings, with only few exceptions. For instance, the Dehn filling of a Seifert manifold is again a Seifert manifold, with only one exception (a fibre-parallel Dehn filling gives a connected sum of Seifert manifolds, see Corollary 10.3.44).

The most striking appearance of this phenomenon is the Hyperbolic Dehn filling Theorem, proved by Thurston at the end of the 1970s. This theorem says roughly that "most" Dehn fillings of a cusped hyperbolic manifold are still hyperbolic.

15.1.1. Generalised Dehn filling parameters. In all this chapter, we consider a compact oriented three-manifold N with $\partial N = T_1 \sqcup \ldots \sqcup T_c$ consisting of tori, and we fix generators m_i, l_i for every $\pi_1(T_i)$.

Let a *generalised Dehn filling parameter* $s = (s_1, \ldots, s_c)$ be a sequence where each s_i is either the symbol ∞ or a rationally related pair of real numbers $(p, q) = (kp', kq') = k(p', q')$ where $k > 0$ is real and (p', q') are coprime integers. We think at s_i as lying in the two-sphere $S^2 = \mathbb{R}^2 \cup \{\infty\}$.

A generalised Dehn filling parameter s determines a Dehn filling N^{fill} of N as follows: for every i, if $s_i = k(p', q')$ we fill T_i by killing the slope $p'm_i + q'l_i$, while if $s_i = \infty$ we do nothing. The result is a new compact manifold, that may be closed (if there are no ∞ in s) or be bounded by some tori. We also mark the cores of the filled solid tori with the label $\alpha_i = \frac{2\pi}{k} > 0$.

15.1.2. The Hyperbolic Dehn filling Theorem. This chapter is mainly devoted to the proof of the following theorem.

Theorem 15.1.1. *Let $M = \text{int}(N)$ be a complete orientable finite-volume cusped hyperbolic three-manifold. There is a neighbourhood U of $(\infty, \dots, \infty)$ in $S^2 \times \dots \times S^2$ such that for every generalised Dehn filling parameter $s \in U$ the interior M^{fill} of the manifold N^{fill} obtained by Dehn filling N along s admits a finite-volume complete hyperbolic structure with cone angles.*

The cores of the filling solid tori are closed geodesics with cone angles α_i. The singular locus of M^{fill} consists of the core geodesics with $\alpha_i \neq 2\pi$.

If $s \in U$ is an ordinary (not extended) Dehn filling parameter (see Section 14.2.1), that is if s_i is either ∞ or a coprime pair (p, q) for all i, then $\alpha_i = 2\pi$ and N^{fill} is a hyperbolic manifold without cone angles: this is of course the case of most interest.

Corollary 15.1.2. *Let $M = \text{int}(N)$ be a complete orientable finite-volume hyperbolic three-manifold. For every $i = 1, \dots, c$ there is a finite set S_i of slopes in T_i such that for every Dehn filling parameter s with $s_i \notin S_i$ for all i, the filled manifold $M^{\text{fill}} = \text{int}(N^{\text{fill}})$ is hyperbolic.*

When M has one cusp, the corollary may be stated simply as follows.

Corollary 15.1.3. *If M is a complete orientable finite-volume hyperbolic three-manifold with one cusp, all but finitely many Dehn fillings M^{fill} are hyperbolic.*

To appreciate the power of these theorems, consider for instance a hyperbolic link $L \subset S^3$ (recall that $L \subset S^3$ is hyperbolic if $S^3 \setminus L$ admits a finite-volume complete hyperbolic metric). By the hyperbolic Dehn filling theorem, there is a finite subset $S \subset \mathbb{Q}$ such that every surgery on L with coefficients in $\mathbb{Q} \setminus S$ produces a closed hyperbolic manifold.

15.1.3. Examples. We describe a couple of clarifying examples. We know that the figure eight knot $K \subset S^3$ is hyperbolic: by the Dehn filling Theorem, there is an open neighbourhood $U \subset S^2$ of ∞ such that every Dehn surgery on K with parameter $s \in U$ produces a hyperbolic cone manifold.

We will prove in the next section that the subset U shown in Figure 15.1 fulfils this requirement. The only coprime pairs (p, q) that are not
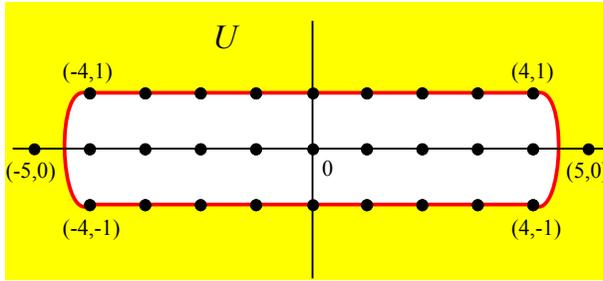
Figure 15.1. Hyperbolic Dehn fillings on the figure-eight knot complement.

$\frac{p}{q}$	manifold
0	$T_{\left(\begin{smallmatrix} 3 & 1 \\ -1 & 0 \end{smallmatrix}\right)}$
±1	$\left(S^2, (2,1), (3,1), (7,-6)\right)$
±2	$\left(S^2, (2,1), (4,1), (5,-4)\right)$
±3	$\left(S^2, (3,1), (3,1), (4,-3)\right)$
±4	$\left(D, (2,1), (2,1)\right) \bigcup_{\left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)} \left(D, (2,1), (3,1)\right)$

Table 15.1. The non-hyperbolic Dehn surgeries of the figure-eight knot. We get a torus bundle with Anosov monodromy and hence of Sol geometry, three Seifert manifolds of $\widetilde{\mathrm{SL}}_2$ geometry, and a graph manifold that splits into two Seifert manifolds via a map that interchanges the fibres and the boundary sections of the two portions (the map is here denoted by a matrix in the section-fibre basis).

contained in U are those with $\frac{p}{q}$ equal to one of the following

(21) $\qquad\qquad -4, -3, -2, -1, 0, 1, 2, 3, 4, \infty.$

Every other surgery on K yields a closed hyperbolic three-manifold. For instance, the manifold Vol2 from Table 14.5 can be obtained with $\frac{p}{q} = \pm 5$ (there is a symmetry of K sending $\frac{p}{q}$ to $-\frac{p}{q}$).

On the other hand, a $\frac{p}{q}$ surgery from the list (21) does not produce a hyperbolic manifold: see Table 15.1, taken from Martelli – Petronio [**38**].

More generally, a parameter $k(p, q) \in U$ produces a hyperbolic closed manifold with a core geodesic of cone angle $\frac{2\pi}{k}$, which may be interpreted as an orbifold when $k \in \mathbb{N}$. In particular, if $(p, q) = (1, 0)$ the closed manifold is just S^3, because an ∞-surgery on a knot in S^3 gives S^3 back.
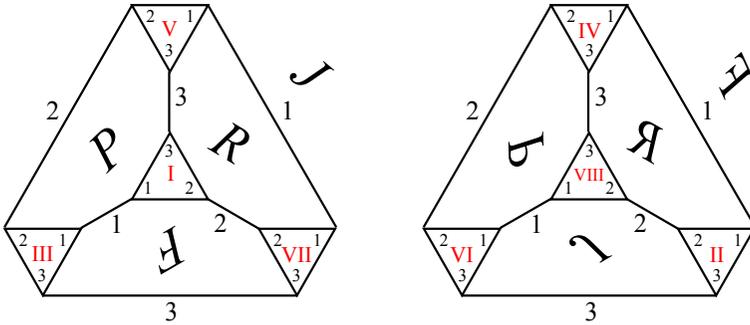
Figure 15.2. We truncate the tetrahedra and flatten their
boundary. We assign a modulus z and w to them. The edge
numbered $i = 1, 2, 3$ has modulus z_i (on the left tetrahedron)
or w_i (on the right tetrahedron).

By looking at Figure 15.1 we discover that for every $k \geqslant 5$ there is a
hyperbolic structure on S^3 with cone angle $\frac{2\pi}{k}$ on the figure-eight knot K.

We now consider a link with two components. The *Whitehead link*
L shown in Figure 14.10 is hyperbolic: it is hard to determine an explicit
open set $U \subset S^2 \times S^2$ in this case, but a careful computer-aided analysis
carried with SnapPy [**15**] shows the following, see [**38**].

Theorem 15.1.4. *A $\left(\frac{p}{q}, \frac{r}{s}\right)$ surgery on the Whitehead link produces a
closed hyperbolic manifold, unless one of the following holds:*

- *either $\frac{p}{q}$ or $\frac{r}{s}$ belongs to the set $\{0, 1, 2, 3, 4, \infty\}$,*
- *up to permutation the pair $\left(\frac{p}{q}, \frac{r}{s}\right)$ belongs to the set*

$$\left\{(-4, -1), (-3, -1), (-2, -2), (-2, -1), \left(\tfrac{3}{2}, 5\right), \left(\tfrac{4}{3}, 5\right) \left(\tfrac{5}{2}, \tfrac{7}{2}\right)\right\}.$$

The proposition is symmetric in the two coefficients because the link
itself has a symmetry that interchanges the two components. All the
Dehn filling parameters listed in the theorem indeed produce non-hyperbolic
manifolds. For instance, if $\frac{p}{q} \in \{1, 2, 3\}$ we get a Seifert manifold fibering
over the disc with two singular fibres, see [**38**].

We first prove Theorem 15.1.1 for the figure-eight knot, where the
combinatorics is so simple that everything can be verified by hand. We will
then prove the theorem in general in Section 15.2 using more sophisticated
tools.

15.1.4. The figure-eight knot example. We now discuss in detail the
standard example: the figure-eight knot complement, one of the very few
cases where the consistency and completeness equations can be solved by
hand. This part will not be needed in the proof of Theorem 15.1.1, so the
reader may wish to skip it and go directly to Section 15.1.5.
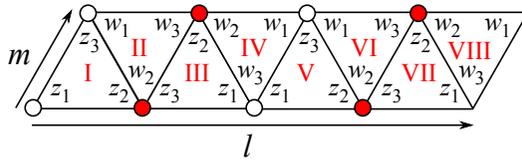
Figure 15.3. The triangulated boundary is a torus obtained by identifying the opposite edges of this parallelogram. The triangulation $\mathcal{T}$ has two edges, that we colour in red and white. Their endpoints are shown here.

Let $\mathcal{T}$ be the triangulation with two tetrahedra shown in Figure 14.7. The truncated version is in Figure 15.2. We assign the complex variable z and w to the left and right tetrahedron, respectively. We set

$$z_1 = z, \quad z_2 = \frac{1}{1-z}, \quad z_3 = \frac{z-1}{z}$$

and define w_1, w_2, w_3 similarly. Recall that

(22) $$z_1 z_2 z_3 = w_1 w_2 w_3 = -1.$$

The modulus of an edge in Figure 15.2 labeled with $i \in \{1, 2, 3\}$ is z_i or w_i depending on the tetrahedron. With some patience one sees that the boundary triangulated surface is a torus T as in Figure 15.3. The triangulation $\mathcal{T}$ has two edges shown in Figure 14.7, each yielding a consistency equation. Figure 15.3 shows that the consistency equations are

$$z_2^2 z_3 w_2^2 w_3 = 1, \quad z_1^2 z_3 w_1^2 w_3 = 1.$$

Using (22) we see that the equations are both equivalent to

$$z_2 w_2 = z_1 w_1,$$

that is

(23) $$z(1-z)w(1-w) = 1.$$

We now look at the boundary torus T. Let m and l be the generators of $\pi_1(T)$ shown in Figure 15.3. We have

(24) $$\mu(m) = -z_1 z_3 w_1 = w(1-z), \quad \mu(l) = z_2^2 z_3^2 w_1^4 w_2^2 w_3^2 = \frac{w^2}{z^2}.$$

The completeness equations are

(25) $$w(1-z) = 1, \quad z^2 = w^2.$$

Exercise 15.1.5. The only solution to (23) and (25) is $z = w = e^{\frac{\pi i}{3}}$.

We have confirmed that the figure-eight complement M has a complete hyperbolic structure obtained by representing both tetrahedra with ideal regular hyperbolic tetrahedra. We now investigate the non-complete solutions of the consistency equations.
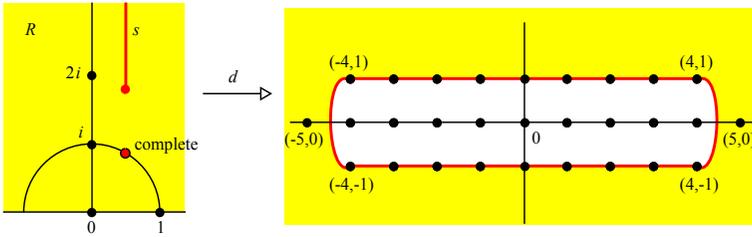
Figure 15.4. Every $w \in R$ determines a hyperbolic structure for M (left). The Dehn filling generalised invariants d map R onto a neighbourhood of ∞ (right).

The two ideal tetrahedra have moduli w and z and the consistency equations reduce to one equation (23) which we rewrite as

$$z^2 - z + \frac{1}{w(1-w)} = 0.$$

The solutions are

$$z = \frac{1 \pm \sqrt{1 - \frac{4}{w(1-w)}}}{2}.$$

We are only interested in solutions with $\Im z, \Im w > 0$. For every w with $\Im w > 0$ there is a unique solution z with $\Im z > 0$, except when $\Delta \in \mathbb{R}_{\geqslant 0}$.

Exercise 15.1.6. We have $\Delta = 1 - \frac{4}{w(1-w)} \in \mathbb{R}_{\geqslant 0}$ if and only if w belongs to the half-line $s = \left\{ \frac{1}{2} + yi \text{ with } y \geqslant \frac{\sqrt{15}}{2} \right\}$, see Figure 15.4-(left).

We define the open region

$$R = \{ w \in \mathbb{C} \mid \Im w > 0 \} \setminus s.$$

Every $w \in R$ determines a hyperbolic structure for M. The complete structure is obtained at $w_0 = \frac{1}{2} + \frac{\sqrt{3}}{2}$, see Figure 15.4-(left).

The manifold M is a knot complement and hence the boundary torus T is equipped with its natural meridian/longitude basis m', l', see Section 11.3.1. Using SnapPea we find out that

$$m' = m, \qquad l' = l + 2m.$$

From (24) we get

(26) $\qquad \tilde{\mu}(m') = w(1-z), \quad \tilde{\mu}(l') = \frac{z^2}{w^2} w^2(1-z)^2 = z^2(1-z)^2.$

We fix m', l' as generators for $\pi_1(T)$. We defined in Section 14.2.7 the generalised Dehn surgery invariant $(p, q) = d(w)$ for every non-complete solution $w \in R \setminus \{w_0\}$. This gives a continuous map

$$d \colon R \longrightarrow S^2$$

that sends w_0 to ∞.

Proposition 15.1.7. *The image $d(R)$ contains the coloured region shown in Figure 15.4-(right).*

Proof. The domain R is an open disc and its abstract closure $\bar{R}$ is homeomorphic to the closed disc. We show that d extends to a continuous map $\bar{R} \to S^2$ which sends ∂R to the almost-rectangle shown in Figure 15.4.

The invariants $d(w) = (p, q)$ are such that

(27) $$w^p(1 - z)^{p+2q} z^{2q} = e^{2\pi i}.$$

The region R has two involutions (determined in fact by isometries of M):

- the involution $\tau(w) = z = \frac{1 \pm \sqrt{1 - \frac{4}{w(1-w)}}}{2}$. This involution sends $\mu(m)$ to $\mu(m)^{-1} = \mu(m^{-1})$, hence $d(\tau(w)) = -d(w)$;
- the involution $\sigma(w) = \overline{1 - w}$: using (26) we get $d(\sigma(w)) = \overline{d(w)}$.

Now we note that when w tends to a point in the line $l_+ = [1, +\infty]$ then z tends to a point in $l_- = [-\infty, 0]$. Therefore the argument of $w, z, 1 - z$ tends respectively to $0, \pi, 0$. Equation (27) implies that $q \to 1$.

If $w \to 1$ then $z \to -\infty$ and (27) implies that $p + 2q + 2q = p + 4q \to 0$, so $(p, q) \to (-4, 1)$. If $w \to +\infty$ then $z \to 0$ and equation $z(1 - z)w(1 - w) = 1$ implies that $|z||w|^2 \to 1$; here (27) gives $|w|^{p-4q} \to 1$, hence $p - 4q \to 0$, so $(p, q) \to (4, 1)$. We have proved that d maps $[1, +\infty]$ onto the segment

$$\overline{(-4, +1), (+4, +1)}.$$

Using the involution τ we deduce that d maps $[-\infty, 0]$ onto the segment

$$\overline{(+4, -1), (-4, -1)}.$$

When w is near the right side of the half-line s, the number z tends to the segment $(0, \frac{1}{2}]$. Therefore the arguments of z and $1 - z$ tend to 0 and (27) implies that the argument of w^p tends to 2π. When $w \in s$ the argument of w is at least $\arctan \sqrt{15} = 1.31811607\ldots$ and hence p is at most $\frac{2\pi}{\arctan \sqrt{15}} = 4.374 \ldots < 5$. This implies that we can connect $(4, 1)$ and $(4, -1)$ while staying inside the image of d with a curve as in Figure 15.4. $\qquad \square$

15.1.5. A road map. The rest of this chapter is mostly devoted to the proof of Theorem 15.1.1. Here is our plan: we first study the combinatorial properties of an arbitrary ideal triangulation $\mathcal{T}$ for M, and we prove that the solutions to the consistency equations form a complex manifold $\mathrm{Def}(M, \mathcal{T})$ of (complex) dimension c equal to the number of cusps of M. The manifold $\mathrm{Def}(M, \mathcal{T})$ is nicely parametrized by the holonomies on any c fixed curves at different cusps.

The manifold $\mathrm{Def}(M, \mathcal{T})$ is often empty, but since M is hyperbolic we expect that there should be some ideal triangulation $\mathcal{T}$ for which $\mathrm{Def}(M, \mathcal{T})$

is not empty and contains a solution z that also satisfies the completeness equations (we are unfortunately not able to prove this in general, and this will be a technical issue). The generalised Dehn filling invariants defined in Section 14.2.7 furnish a local diffeomorphism $d \colon \mathrm{Def}(M, \mathcal{T}) \to S^2 \times \ldots \times S^2$ that sends z to $(\infty, \ldots, \infty)$. In particular this map is open and its image covers a neighbourhood U of $(\infty, \ldots, \infty)$. We conclude thanks to the discussion of Section 14.2.7.

We start by defining and exploring the solution space $\mathrm{Def}(M, \mathcal{T})$.

15.2. The solution space

In this section we study the combinatorial properties of ideal triangulations of three-manifolds in general. We study the space of solutions to the consistency equations, we prove that it is a complex manifold and exhibit a concrete parametrisation via the holonomy of peripheral curves.

15.2.1. Edges and tetrahedra. Throughout this section, we let N be a compact oriented three-manifold N with $\partial N = T_1 \sqcup \ldots \sqcup T_c$ consisting of tori, and define $M = \mathrm{int}(N)$. Let $\mathcal{T}$ be any ideal triangulation for M.

Proposition 15.2.1. *The ideal triangulation $\mathcal{T}$ has the same number n of tetrahedra and edges.*

Proof. The total space $|\mathcal{T}|$ of the triangulation $\mathcal{T}$ has v vertices, e edges, f faces, and t tetrahedra. Since $\chi(\partial N) = 2\chi(N)$ for every compact 3-manifold N and ∂N consists of tori, we get $\chi(N) = 0$. Therefore

$$v = v + \chi(N) = \chi(|\mathcal{T}|) = v - e + f - t = v - e + t$$

since $f = 2t$. Then $e = t$. $\qquad\square$

Recall that in the consistency equations we get one complex variable z_i for each tetrahedron and one equation for each edge; since we have the same number of edges and tetrahedra, one would guess that the solutions form typically a discrete set of points, but this is surprisingly not the case: we now show that some equations are redundant, leaving some space for a higher-dimensional space of solutions. The origin of this redundancy lies in the combinatorial properties of three-dimensional triangulations.

15.2.2. Incidence matrices. We denote respectively by $\Delta_1, \ldots, \Delta_n$ and $e_1, \ldots, e_n$ the tetrahedra and edges of $\mathcal{T}$. We fix an orientation on M, which induces an orientation on each tetrahedron, and we assign numbers 1, 2, and 3 to pairs of opposite edges on each Δ_i in a way that is orientation-preservingly isomorphic to Figure 15.5.

We now define a $2n \times 3n$ integral matrix A that encodes some combinatorial adjacencies between tetrahedra. The matrix A should be seen as a 2×3 rectangle of $n \times n$ sub-matrices as in Figure 15.6-(left), where:
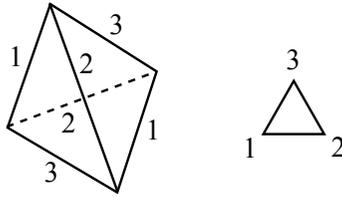
Figure 15.5. We label the edges of every *oriented* tetrahedron as shown: the labelling is determined once we chose arbitrarily a pair of opposite edges and label them with 1 (left). The vertices of every *oriented* triangle in the triangulated boundary inherit labels as in the figure (right).
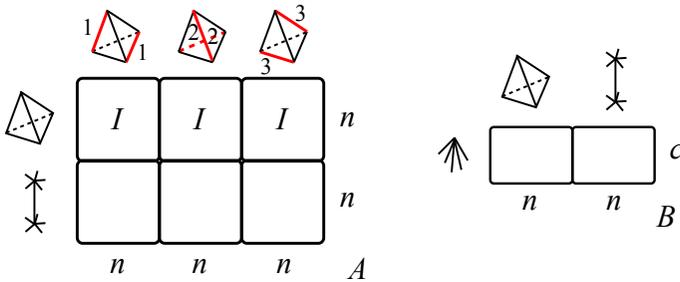
Figure 15.6. The incidence matrices A and B count the incidences between edge pairs, tetrahedra, edges, and ideal vertices.

- each column of A corresponds to a pair of opposite edges of one tetrahedron, more precisely the columns $i, i+n$, and $i+2n$ correspond the opposite edges of type 1,2,3 of Δ_i;
- the first n rows of A correspond to the tetrahedra $\Delta_1, \ldots, \Delta_n$;
- the last n rows of A correspond to the edges $e_1, \ldots, e_n$.

When $i \leqslant n$, the entry A_{ij} is 1 or 0 depending on whether Δ_i contains the j-th edge pair. In other words, the top three $n \times n$ sub-matrices of A are three identity matrices, see Figure 15.6-(left).

When $i = n + i'$ with $i' > 0$, the entry A_{ij} counts how many edges in the j-th edge pair are glued to the i'-th edge $e_{i'}$: this number is hence 0, 1, or 2.

Exercise 15.2.2. The entries on each column of A sum to 3.

Let now $v_1, \ldots, v_c$ be the ideal vertices of $\mathcal{T}$. We define another $c \times 2n$ integral matrix B, which consists of two $c \times n$ sub-matrices as in Figure 15.6-(right), where:

- the rows of B correspond to the ideal vertices $v_1, \ldots, v_c$;

- the first n columns of B correspond to the n tetrahedra $\Delta_1, \dots, \Delta_n$;
- the last n columns correspond to the edges $e_1, \dots, e_n$.

When $j \leqslant n$, the entry B_{ij} is *minus* the number of times Δ_j is incident to v_i, and when $j = n + j'$ with $j' > 0$ the entry B_{ij} is *plus* the number of times $e_{j'}$ is incident to v_i. The possible entries are $0, -1, -2, -3, -4$ and $0, 1, 2$ respectively.

Exercise 15.2.3. The entries on each column of B sum to -4 or 2.

The matrices A and B are designed to be interpreted as linear maps, and to be composed.

Proposition 15.2.4. *The following sequence is exact:*

$$\mathbb{R}^{3n} \xrightarrow{A} \mathbb{R}^{2n} \xrightarrow{B} \mathbb{R}^c \longrightarrow 0.$$

The proof of this proposition splits into three lemmas.

Lemma 15.2.5. *We have $BA = 0$.*

Proof. The i-th row of B corresponds to the ideal vertex v_i and the k-th column of A corresponds to the k-th pair of opposite edges. The sum $\sum_{j=1}^{2n} B_{ij} A_{jk}$ counts *minus* the number of times v_i is incident to the tetrahedron containing the edge pair, *plus* the number of times it is incident to the two edges of the pair. These numbers are equal and hence we get 0. $\qquad\square$

Lemma 15.2.6. *The c rows in B are independent vectors.*

Proof. Suppose there is a vanishing linear combination $\sum_{i=1}^c \lambda_i B_i = 0$ of the rows B_i of B. If the edge e_j has endpoints at v_a and v_b, the entries in the $(n+j)$-th column B^{n+j} are 1 at the rows a and b and zero everywhere else: therefore we get $\lambda_a + \lambda_b = 0$. If we apply this argument to the three edges of a triangle we get

$$\lambda_a + \lambda_b = 0, \quad \lambda_b + \lambda_c = 0, \quad \lambda_c + \lambda_a = 0$$

which implies $\lambda_a = \lambda_b = \lambda_c = 0$. Then all coefficients λ_a vanish. $\qquad\square$

Lemma 15.2.7. *The c rows in B generate $\ker(A^{\mathsf{T}})$.*

Proof. Since $BA = 0$ implies $A^{\mathsf{T}} B^{\mathsf{T}} = 0$, we already know that the rows in B are contained in $\ker(A^{\mathsf{T}})$, and we must prove that they generate it. We pick a generic horizontal vector $q = (q_1, \dots, q_{2n}) \in \mathbb{R}^{2n}$ with $qA = 0$ and we need to prove that q is generated by the rows B_i of B.

Recall that the columns A^j, A^{j+n}, A^{j+2n} of A correspond to the edge pairs of type 1, 2, and 3 of the tetrahedron Δ_j. We label the vertices of Δ_j as a, b, c, d, the edges of Δ_j as pairs like ab, and the corresponding vertices and edges in $\mathcal{T}$ as $j(a)$ and $j'(ab)$, interpreted as numbers in $\{1, \dots, c\}$ and $\{1, \dots, n\}$ respectively. Moreover we set $j(ab) = j'(ab) + n$. We suppose that ab, ac, ad are of type 1, 2, 3.

The conditions $qA^j = qA^{j+n} = qA^{j+2n} = 0$ say that

$$qA^j = q_j + q_{j(ab)} + q_{j(cd)} = 0,$$
$$qA^{j+n} = q_j + q_{j(ac)} + q_{j(bd)} = 0,$$
$$qA^{j+2n} = q_j + q_{j(ad)} + q_{j(bc)} = 0.$$

This implies that

(28) $$q_{j(ab)} + q_{j(cd)} = q_{j(ac)} + q_{j(bd)} = q_{j(ad)} + q_{j(bc)} = -q_j.$$

We now prove that $q = \sum \lambda_i B_i$ where the coefficients λ_i are defined as follows. Consider an ideal triangle in Δ_j with vertices a, b, c and define

$$\lambda_{j(a)} = \frac{q_{j(ab)} + q_{j(ca)} - q_{j(bc)}}{2}.$$

We check that $\lambda_{j(a)}$ does not depend on the ideal triangle: the triangle in Δ_j with vertices a, b, d gives

$$\lambda_{j(a)} = \frac{q_{j(ab)} + q_{j(da)} - q_{j(bd)}}{2} = \frac{q_{j(ab)} + q_{j(ca)} - q_{j(bc)}}{2}$$

using (28). Two triangles incident to the ideal vertex $v_{j(a)}$ are connected by a path of adjacent tetrahedra and hence $\lambda_{j(a)}$ is well-defined.

Finally, we need to check that indeed $q = \sum \lambda_i B_i$. We consider the edge $j(ab)$ and recall that the $j(ab)$-th column of B is everywhere zero except at rows $j(a)$ and $j(b)$; hence the $j(ab)$-th entry of $\sum \lambda_i B_i$ is $\lambda_{j(a)} + \lambda_{j(b)}$, which by definition equals $q_{j(ab)}$, so we are done.

Analogously, the j-th column of B is everywhere zero except at the rows $j(a), j(b), j(c), j(d)$ and hence the j-th entry of $\sum \lambda_i B_i$ is $-(\lambda_{j(a)} + \lambda_{j(b)} + \lambda_{j(c)} + \lambda_{j(d)})$ which equals $-q_{j(ab)} - q_{j(cd)} = q_j$. The proof is complete.   □

These three lemmas imply Proposition 15.2.4. In particular we get the following:

Corollary 15.2.8. *We have* $\mathrm{rk} A = 2n - c$.

15.2.3. A symplectic form. We define an alternating bilinear form ω on $\mathbb{R}^{3n}$ as follows: for every $i = 1, \ldots, n$ consider the form

$$\omega_i = \begin{pmatrix} 0 & 1 & -1 \\ -1 & 0 & 1 \\ 1 & -1 & 0 \end{pmatrix}$$

on the subspace V_i generated by e_i, e_{i+n}, and e_{i+2n}, and define $\omega = \oplus_i \omega_i$. Strictly speaking, the form ω is not symplectic because it is degenerate: its radical is generated by all vectors of type $e_i + e_{i+n} + e_{i+2n}$, that is by the first n rows of the matrix A. The form ω becomes symplectic after quotienting $\mathbb{R}^{3n}$ by the radical.

Proposition 15.2.9. *The rows of A generate a lagrangian subspace.*

Proof. In other words, we need to prove that $\omega(A_j, A_k) = 0$ for any rows A_j, A_k of A. The first n rows lie in the radical, hence we are left to check that $\omega(A_{n+j}, A_{n+k}) = 0$ for all $0 < j < k \leqslant n$ corresponding to the j-th and k-th edge of the triangulation.

Each tetrahedron Δ_i contributes to the number $\omega(A_{n+j}, A_{n+k})$ as follows: if two edges contained in some face f of Δ_i correspond to the edges j and k of the triangulation, they contribute with a ± 1; the face f is also contained in an adjacent tetrahedron where they contribute with opposite sign, and hence everything sums to zero.   □

15.2.4. The solution space. We now employ the combinatorial results of the previous sections to study the consistency equations for $\mathcal{T}$. We assign the variables z_j, $\frac{1}{1-z_j}$, and $\frac{z_j-1}{z_j}$ to the edges of type 1, 2, and 3 in each tetrahedron Δ_j. These three variables lie in the upper half-plane

$$\mathbb{C}_+ = \{\Im z > 0\}$$

which we see as the subset $\mathbb{C}_+ \subset \widetilde{\mathbb{C}^*}$ consisting of all $\rho e^{i\theta}$ with $0 < \theta < \pi$. The consistency equation at the edge e_i is the following equation in $\widetilde{\mathbb{C}^*}$:

$$(29) \qquad g_i(z) = \prod_{j=1}^{n} z_j^{A_{n+i,j}} \left(\frac{1}{1-z_j}\right)^{A_{n+i,n+j}} \left(\frac{z_j-1}{z_j}\right)^{A_{n+i,2n+j}} = e^{2\pi i}.$$

The elements $\frac{1}{1-z_j}, \frac{z_j-1}{z_j} \in \mathbb{C}_+ \subset \widetilde{\mathbb{C}^*}$ depend holomorphically on $z_j \in \mathbb{C}_+$. We have defined a complex function

$$g \colon \mathbb{C}_+^n \longrightarrow \left(\widetilde{\mathbb{C}^*}\right)^n$$

by setting $g = (g_1, \ldots, g_n)$. The *solution space* is the set

$$\mathrm{Def}(M, \mathcal{T}) = g^{-1}(e^{2\pi i}, \ldots, e^{2\pi i}).$$

This space consists precisely of the solutions to the consistency equations of $\mathcal{T}$. Each solution $z \in \mathrm{Def}(M, \mathcal{T})$ gives a hyperbolic metric on M, which "deforms" as z varies: this justifies the use of the symbol "Def". We want to prove that $\mathrm{Def}(M, \mathcal{T})$ is a complex manifold of dimension c.

15.2.5. Complex differentials. We now use the biholomorphic map $\log \colon \widetilde{\mathbb{C}^*} \to \mathbb{C}$ and define the function $G(z) = \log g(z)$ for all $z \in \mathbb{C}_+^n$. Now

$$\mathrm{Def}(M, \mathcal{T}) = G^{-1}(2\pi i, \ldots, 2\pi i).$$

We may write

$$G_i(z) = \sum_{j=1}^{n} A'_{i,j} \log z_j + A'_{i,n+j} \log(1-z_j) + N_i \pi i$$

where A' is the $n \times 2n$ integral matrix defined by

$$A'_{i,j} = A_{n+i,j} - A_{n+i,2n+j}, \quad A'_{i,n+j} = -A_{n+i,n+j} + A_{n+i,2n+j}$$

and N_i is some fixed integer that depends on the way we interpret $\log(1 - z_j)$. The complex differential at z splits as a composition of linear maps

$$dG_z : \mathbb{C}^n \xrightarrow{Z} \mathbb{C}^{2n} \xrightarrow{A'} \mathbb{C}^n$$

where

$$Z = \begin{pmatrix} \frac{1}{z_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{z_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{z_n} \\ \frac{1}{z_1-1} & 0 & \cdots & 0 \\ 0 & \frac{1}{z_2-1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{z_n-1} \end{pmatrix}$$

is obtained by differentiating $\log z_j$ and $\log(1 - z_j)$. Let now

$$B' : \mathbb{C}^n \longrightarrow \mathbb{C}^c$$

be the linear map defined by the right $(c \times n)$-sub-matrix B' of B, that is

$$B'_{i,j} = B_{i,n+j}.$$

Proposition 15.2.10. *The following sequence is exact:*

$$\mathbb{C}^{2n} \xrightarrow{A'} \mathbb{C}^n \xrightarrow{B'} \mathbb{C}^c \longrightarrow 0.$$

Proof. Lemma 15.2.6 shows that B is surjective, and its proof actually shows that B' is surjective. Recall that A decomposes into six square matrices, and via Gauss moves on columns we get

$$(30) \quad A = \begin{pmatrix} I & I & I \\ A^1 & A^2 & A^3 \end{pmatrix} \longrightarrow \begin{pmatrix} 0 & 0 & I \\ A^1 - A^3 & -A^2 + A^3 & A^3 \end{pmatrix} = \begin{pmatrix} 0 & I \\ A' & A^3 \end{pmatrix}$$

Now $BA = 0 \Rightarrow B'A' = 0$, and $\mathrm{rk}\,A = 2n - c \Rightarrow \mathrm{rk}\,A' = n - c$. □

This shows in particular that the image of G lies in a fixed affine subspace parallel to $\ker B'$ of complex codimension c. We now consider the standard symplectic form $\omega = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}$ on $\mathbb{R}^{2n}$.

Proposition 15.2.11. *The rows of A' generate a lagrangian subspace.*

Proof. We know that the rows of A generate a lagrangian subspace with respect to the degenerate form on $\mathbb{R}^{3n}$ by Proposition 15.2.9. The Gauss moves (30) imply the assertion. □

Proposition 15.2.12. *For every $z \in \mathbb{C}^n_+$ the following sequence is exact:*

$$\mathbb{C}^n \xrightarrow{dG_z} \mathbb{C}^n \xrightarrow{B'} \mathbb{C}^c \longrightarrow 0.$$

Proof. We have $dG_z = A'Z$. Proposition 15.2.10 gives an exact sequence

$$\mathbb{C}^{2n} \xrightarrow{A'} \mathbb{C}^n \xrightarrow{B'} \mathbb{C}^c \longrightarrow 0.$$

We only need to prove that $\mathrm{Im}\,(A'Z) = \mathrm{Im}\,A'$, that is $\mathrm{rk}(A'Z) = n - c$. We prefer to consider the transpose matrix $M = (A'Z)^{\mathsf{T}} = Z^{\mathsf{T}}(A')^{\mathsf{T}}$. Then

$$Z^{\mathsf{T}} = \begin{pmatrix} \frac{1}{z_1} & 0 & \cdots & 0 & \frac{1}{z_1-1} & 0 & \cdots & 0 \\ 0 & \frac{1}{z_2} & \cdots & 0 & 0 & \frac{1}{z_2-1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{z_n} & 0 & 0 & \cdots & \frac{1}{z_n-1} \end{pmatrix}.$$

The map $Z^{\mathsf{T}} \colon \mathbb{C}^{2n} \to \mathbb{C}^n$ is obviously not injective, but its restriction

$$Z^{\mathsf{T}}|_{\mathbb{R}^{2n}} \colon \mathbb{R}^{2n} \longrightarrow \mathbb{C}^n$$

is an isomorphism of real spaces because $\frac{1}{z_i}$ and $\frac{1}{z_i-1}$ are independent over $\mathbb{R}$ (we use here that $z_i \notin \mathbb{R}$).

We define a symplectic form Ω on the real vector space $\mathbb{C}^n$ by pushing-forward ω along $Z^{\mathsf{T}}|_{\mathbb{R}^{2n}}$, that is we set $\Omega(Z^{\mathsf{T}}v, Z^{\mathsf{T}}w) = \omega(v, w)$. By construction Ω is a sum of symplectic forms on the coordinate complex lines of $\mathbb{C}^n$. Note that a symplectic form on $\mathbb{C}$ is unique up to rescaling.

Since $\mathrm{Im}\,z_i > 0$, the complex numbers $\frac{1}{z_i}$ and $\frac{1}{z_i-1}$ form a negatively-oriented $\mathbb{R}$-basis of $\mathbb{C}$, and this implies that $\Omega(v, iv) < 0$ for every non-zero vector $v \in \mathbb{C}^n$. In particular $\langle \cdot, \cdot \rangle = -\Omega(\cdot, i\cdot)$ is a positive-definite scalar product that induces a norm $\| \cdot \|$ on $\mathbb{C}^n$.

On each coordinate complex line of $\mathbb{C}^n$, the form Ω is just a rescaling of the standard symplectic form and hence $\langle , \rangle$ is a rescaling of the standard scalar product. In particular $\Omega(iv, iw) = \Omega(v, w)$ and $\langle iv, iw \rangle = \langle v, w \rangle$.

Finally, we suppose by contradiction that $\mathrm{rk}M < n - c$. Since $Z^{\mathsf{T}}|_{\mathbb{R}^{2n}}$ is injective, the real subspace $M(\mathbb{R}^n)$ has real dimension $n - c$. Since $\mathrm{rk}M < n-c$, there are vectors $v_1, v_2 \in \mathbb{R}^n$ such that $M(v_1 + iv_2) = 0$ while $Mv_1 \neq 0$ and $Mv_2 \neq 0$. Proposition 15.2.11 gives

$$\Omega(Mv_1, Mv_2) = \omega((A')^{\mathsf{T}}v_1, (A')^{\mathsf{T}}v_2) = 0$$

and therefore

$$\begin{aligned} 0 = \|M(v_1 + iv_2)\|^2 &= \|Mv_1\|^2 + \|iMv_2\|^2 + 2\langle Mv_1, iMv_2 \rangle \\ &= \|Mv_1\|^2 + \|Mv_2\|^2 - 2\Omega(Mv_1, -Mv_2) \\ &= \|Mv_1\|^2 + \|Mv_2\|^2 \end{aligned}$$

implies $Mv_1 = Mv_2 = 0$, a contradiction. $\qquad\square$

Corollary 15.2.13. *The solution space* $\mathrm{Def}(M, \mathcal{T}) \subset \mathbb{C}^n_+$ *is either empty or a complex submanifold of dimension* c.
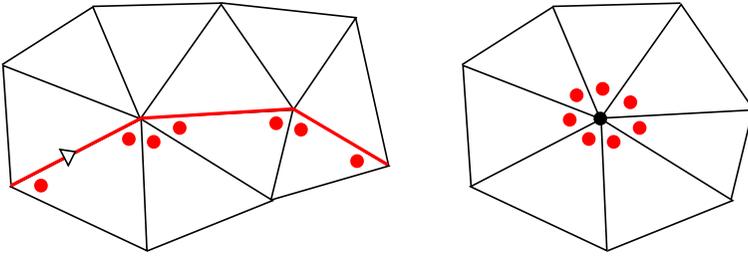
Figure 15.7. We mark with a dot all the interior angles that γ encounters on its right as in the picture: each dot determines an edge in some tetrahedron Δ_j and hence an edge pair (left). We also consider the "degenerate path" consisting of a single point "run counterclockwise", and recover the last n rows of A in this way (right).

Proof. The image of $G\colon \mathbb{C}^n_+ \to \mathbb{C}^n$ lies entirely in a $(n-c)$-dimensional complex affine space S parallel to $\ker B' \subset \mathbb{C}^n$ and by the previous proposition the holomorphic map $G\colon \mathbb{C}^n_+ \to S$ is a submersion onto its image. Therefore the counterimage of any point is a either empty or is a complex submanifold of dimension $n - (n - c) = c$. $\square$

We now want to define a nice holomorphic parametrisation for the manifold $\mathrm{Def}(M, \mathcal{T})$, and to this purpose we need to investigate further the combinatorial properties of $\mathcal{T}$ by looking more closely at its ideal vertices.

15.2.6. Boundary curves. Recall that $M = \mathrm{int}(N)$ and the ideal vertices $v_1, \ldots, v_c$ of $\mathcal{T}$ correspond to the boundary tori $T_1, \ldots, T_c$ of N, which are triangulated by $\mathcal{T}$. Recall also that there are n tetrahedra $\Delta_1, \ldots, \Delta_n$ in $\mathcal{T}$ and each Δ_i has three edge pairs labeled with $i, n + i$, and $2n + i$.

Every simplicial closed oriented curve γ in a triangulated torus T_i determines an integer vector $v_\gamma \in \mathbb{R}^{3n}$ that counts the (right-)incidences between γ and the $3n$ edge pairs of $\mathcal{T}$. More precisely, each interior angle of a triangle in T_i determines an edge in some tetrahedron Δ_j and hence an edge pair: we define $(v_\gamma)_j$ to be the number of times γ encounters the j-th edge pair on its right, as in Figure 15.7-(left).

We defined an alternating degenerate form ω on $\mathbb{R}^{3n}$ and we now want to extend Proposition 15.2.9. We denote by $\gamma \cdot \gamma'$ the algebraic intersection of γ and γ' in $H_1(\partial M, \mathbb{Z})$.

Proposition 15.2.14. *Let γ, γ' be simplicial closed curves in ∂N. Then*

- $\omega(v_\gamma, v_{\gamma'}) = 2\gamma \cdot \gamma'$,
- $\omega(v_\gamma, A_i) = 0$ *for every row A_i of A.*

Proof. The row A_i with $i \leqslant n$ lies in the radical of ω, and A_{n+i} records the incidences of the i-th edge e_i of $\mathcal{T}$ with the $3n$ edge pairs. We may
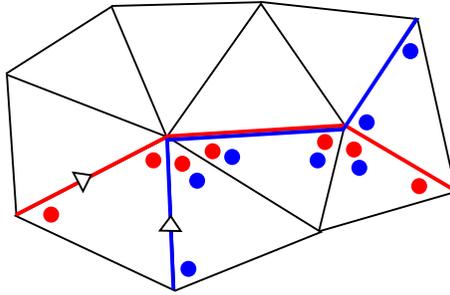
Figure 15.8. We draw the paths γ, γ' and their dots in blue
and red.
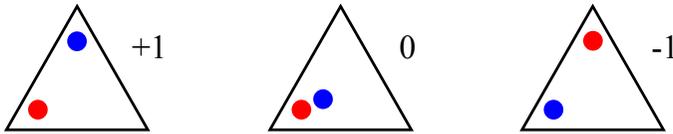
Figure 15.9. Two dots in a triangle contribute to $\omega(\gamma, \gamma')$
with $+1$, 0, or -1 according to their mutual position.

write $A_{n+i} = v_{\gamma'}$ where γ' is a "degenerate" closed path consisting of a
single point (one of the endpoints of e_i), see Fig 15.7-(right). Now the
second assertion is just a special case of the first.

We draw the two paths γ and γ' in blue and red as in Figure 15.8,
marking with dots the angles they encounter on their right. A couple of
blue and red dots in the same triangle contribute to $\omega(v_\gamma, v_{\gamma'})$ according
to their mutual position as shown in Figure 15.9. Therefore portions of
blue and red paths intersecting the same triangle contribute to $\omega(v_\gamma, v_{\gamma'})$
as shown in Figure 15.10.

For every contribution $+1$ as in Figure 15.10-(top-left), there is an
analogous but opposite contribution -1 provided by the other triangle
adjacent to the edge joining the coloured endpoints: these all cancel. We
are left with a $+1$ every time the red path exits leftward from an intersection
with the blue path, a -1 every time it enters from the left, and the opposite
contributions with the blue/red colours interchanged. We easily get a total
of $2\gamma \cdot \gamma'$.

The proof is not yet complete, because every tetrahedron Δ_i deter-
mines four triangles in ∂N and some contributions to $\omega(v_\gamma, v_{\gamma'})$ arise also
from paths intersecting distinct triangles of the same Δ_i: we need to prove
that these contributions all cancel.
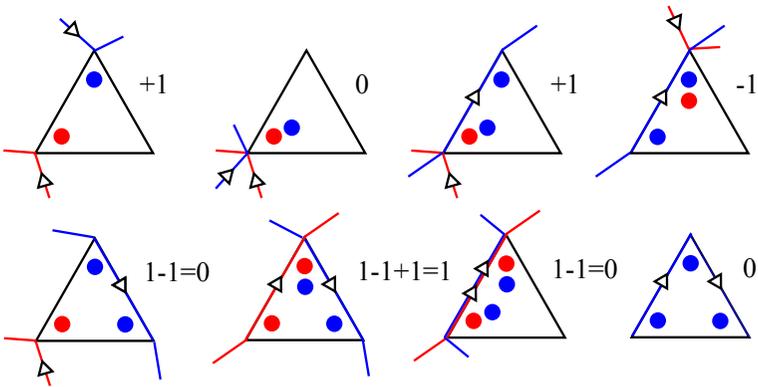
Figure 15.10. Two portions of blue and red paths that intersect the same triangle contribute to $\omega(\gamma, \gamma')$ as shown here. The contribution depends on their mutual configuration, and changes by a sign if we interchange the blue/red colours. If one of the two portions contains at least two edges (as in bottom-right) it determines a vector in the radical, and the contribution is always zero.
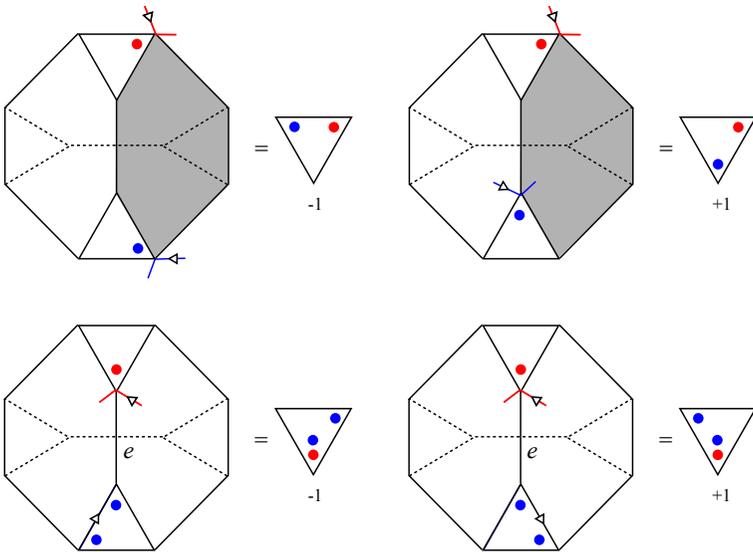
Figure 15.11. When γ and γ' encounter in different truncated triangles of the same tetrahedron Δ_i, they also contribute to $\omega(v_\gamma, v_{\gamma'})$.
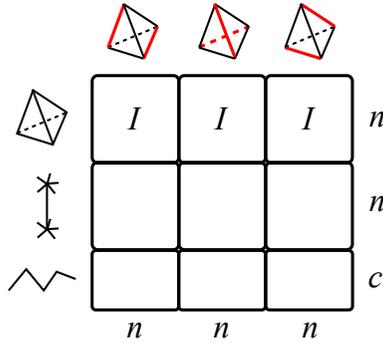
Figure 15.12. The enlarged matrix $\bar{A}$.

Up to symmetries of the tetrahedron and up to interchanging the colours, the possible non-zero contributions are all shown in Figure 15.11. Each configuration as in the first row cancels with an opposite one produced by the tetrahedron adjacent to the grey face. The contributions in the second row arise when γ and γ' cross the endpoints of an edge e on opposite sides, and one checks that the contributions of all the cycle of tetrahedra incident to e sum to zero (exercise). □

15.2.7. Holonomy parameters. We now fix a non-trivial oriented simple closed curve γ_i on each T_i and a representation of γ_i as a simplicial (possibly non-injective) path in the triangulated T_i.

We enlarge the incident $2n \times 3n$ matrix A to a $(2n + c) \times 3n$ matrix $\bar{A}$ by adding the vectors $v_{\gamma_1}, \ldots, v_{\gamma_c}$ at its bottom, see Figure 15.12.

Proposition 15.2.15. *We have* $\mathrm{rk}\bar{A} = 2n$. *The rows of* $\bar{A}$ *generate a maximal lagrangian subspace in* $\mathbb{R}^{3n}$.

Proof. Let $\mu_i \subset T_i$ be a simplicial closed curve with $\gamma_i \cdot \mu_i = 1$. We already know that the first $2n$ rows of $\bar{A}$ generate a lagrangian subspace of dimension $2n - c$, and to conclude it suffices to use Proposition 15.2.14 and deduce that v_{μ_i} is ω-orthogonal to all the rows of $\bar{A}$ except $\bar{A}_{2n+i} = v_{\gamma_i}$. Therefore $\mathrm{rk}\bar{A} = \mathrm{rk}A + c = 2n$. □

Every point $z \in \mathrm{Def}(M, \mathcal{T})$ determines a hyperbolic structure on M and a homomorphism

$$\tilde{\mu}_z \colon \pi_1(T_i) \to \widetilde{\mathbb{C}^*}$$

for each $i = 1, \ldots, c$, see Definition 14.2.1. We define

$$h_i(z) = \tilde{\mu}_z(\gamma_i) = e^{-|\gamma_i|\pi i} \prod_{j=1}^{n} z_j^{\bar{A}_{2n+i,j}} \left(\frac{1}{1 - z_j}\right)^{\bar{A}_{2n+i,n+j}} \left(\frac{z_j - 1}{z_j}\right)^{\bar{A}_{2n+i,2n+j}}$$

and get a holomorphic function

$$h\colon \mathbb{C}_+^n \longrightarrow \left(\widetilde{\mathbb{C}^*}\right)^c.$$

We want to prove that h furnishes a nice parametrisation for the c-dimensional space $\mathrm{Def}(M, \mathcal{T})$. The letter h stands for "holonomy": recall that h_i is the holonomy of γ_i, see Exercise 14.2.6.

As in Section 15.2.5, we define $H(z) = \log h(z)$. We get a map

$$G \times H\colon \mathbb{C}_+^n \longrightarrow \mathbb{C}^n \times \mathbb{C}^c.$$

As above, the differential splits as a composition of linear maps

$$d(G \times H)_z\colon \mathbb{C}^n \xrightarrow{Z} \mathbb{C}^{2n} \xrightarrow{\bar{A}'} \mathbb{C}^n \times \mathbb{C}^c$$

where $\bar{A}'$ is constructed from $\bar{A}$ exactly as A' from A.

Proposition 15.2.16. *The differential $d(G \times H)_z$ is injective $\forall z \in \mathbb{C}_+^n$.*

Proof. As above, Proposition 15.2.15 implies that the rows of $\bar{A}'$ form a lagrangian subspace of dimension n and we conclude with the same proof as Proposition 15.2.12. $\square$

Corollary 15.2.17. *The restriction*

$$h|_{\mathrm{Def}(M,\mathcal{T})}\colon \mathrm{Def}(M, \mathcal{T}) \longrightarrow \left(\widetilde{\mathbb{C}^*}\right)^c$$

is a local biholomorphism.

A *complete solution* $z \in \mathrm{Def}(M, \mathcal{T})$ is one that determines a complete metric on M, that is such that $h(z) = (1, \dots, 1)$.

Corollary 15.2.18. *Every complete solution in $\mathrm{Def}(M, \mathcal{T})$ is isolated.*

The local biholomorphism h depends on the isotopy class of the curves γ_i chosen: different choices give different parametrisations. Note however that the condition that $h_i(z) = 1$ corresponds geometrically to the completeness of T_i and hence does not depend on γ_i.

15.3. Proof of the theorem

We now conclude the proof of Theorem 15.1.1. Through all this section, we let $M = \mathrm{int}(N)$ be a complete orientable finite-volume cusped hyperbolic three-manifold. We have $\partial N = T_1 \cup \dots \cup T_c$ for some $c \geqslant 1$. We also fix a positive pair m_i, l_i of generators for $\pi_1(T_i)$ at each T_i. Recall that each T_i has Euclidean structure, defined up to rescaling.

15.3.1. Near the complete solution. We now suppose that M has a geometric decomposition $\mathcal{T}$ into ideal hyperbolic tetrahedra, in other words the space $\mathrm{Def}(M, \mathcal{T})$ is non-empty and contains a complete solution z^0. We want to study $\mathrm{Def}(M, \mathcal{T})$ near z^0.

We defined in Section 14.2.7 the generalised Dehn filling invariant of a solution $z \in \mathrm{Def}(M, \mathcal{T})$ at T_i to be ∞ if $\mu(m_i) = \mu(l_i) = 1$, or the unique point $(p, q) \in \mathbb{R}^2$ such that $\tilde{\mu}(m_i)^p \tilde{\mu}(l_i)^q = e^{2\pi i}$ otherwise. We denote this invariant by $d_i(z) \in S^2 = \mathbb{R}^2 \cup \{\infty\}$ and get a map

$$d \colon \mathrm{Def}(M, \mathcal{T}) \longrightarrow \underbrace{S^2 \times \ldots \times S^2}_{c}$$

such that $d(z^0) = (\infty, \ldots, \infty)$.

Proposition 15.3.1. *The map d is a local homeomorphism at z^0.*

Proof. We set

$$u_i = \log \tilde{\mu}(m_i), \qquad v_i = \log \tilde{\mu}(l_i).$$

The sets $(u_1, \ldots, u_c)$ and $(v_1, \ldots, v_c)$ are both biholomorphic coordinates on $\mathrm{Def}(M, \mathcal{T})$ which vanish at z^0. Moreover v_i vanishes precisely where u_i does (on solutions that are complete at T_i).

Recall from Exercise 14.2.4 that $\mu = \rho'$ is the derivative of the holonomy ρ at T_i. When $u_i = v_i = 0$ the solution is complete at T_i and up to conjugation we have

$$\rho(m_i) \colon z \mapsto z + 1, \qquad \rho(l_i) \colon z \mapsto z + w_i$$

for some $w_i \in \mathbb{C}$ with $\Im w_i > 0$ which determines the cusp shape of T_i. When $u_i, v_i \neq 0$ the solution is incomplete and

$$\rho(m_i) \colon z \mapsto az, \qquad \rho(l_i) \colon z \mapsto bz$$

for some $a, b \neq 0$. When $u_i, v_i \to 0$, a fundamental domain of $\mathrm{Im}\,\rho$ with vertices at x, ax, bx, abx tends up to rescaling to one with vertices at $x, x + 1, x + w_i, x + 1 + w_i$. Since $a, b \to 1$ we get

$$v_i = \log b \sim b - 1 \sim w_i(a - 1) \sim w_i \log a = w_i u_i$$

where $\sim$ means first-order equality. Therefore $\frac{v_i}{u_i} = w_i$ at the complete solution $v = u = 0$. We deduce in particular that the function $\frac{v_i}{u_i}$ is analytic on $\mathrm{Def}(M, \mathcal{T})$. In what follows we only use that $\Im w_i > 0$.

We consider the following map, defined whenever $(p_i, q_i) \neq (0, 0)$ for all i:

$$\varphi\big((p_1, q_1), \ldots, (p_c, q_c)\big) = 2\pi i \left(\frac{1}{p_1 + q_1 w_1}, \ldots, \frac{1}{p_c + q_c w_c} \right).$$

We extend it continuously to the case $(p_i, q_i) = \infty$ by setting the i-th component of the image to be zero. By composing $\Phi = \varphi \circ d$ we get a map $\Phi \colon \mathrm{Def}(M, \mathcal{T}) \to \mathbb{C}^c$ that sends the complete solution $u = 0$ to

$(0, \ldots, 0)$. We prove that Φ is differentiable and $d\Phi_0$ is invertible: this will conclude the proof. When $u_j, v_j \neq 0$ we have

$$p_j u_j + q_j v_j = 2\pi i$$

and Proposition 14.2.12 gives

$$p_j = -2\pi \frac{\Re v_j}{\Im(\bar{u}_j v_j)}, \qquad q_j = 2\pi \frac{\Re u_j}{\Im(\bar{u}_j v_j)}.$$

Near $u = v = 0$ we have $v_j \sim w_j u_j$ and hence

$$p_j \sim -2\pi \frac{\Re w_j \Re u_j - \Im w_j \Im u_j}{|u_j|^2 \Im w_j}, \qquad q_j \sim 2\pi \frac{\Re u_j}{|u_j|^2 \Im w_j}$$

and therefore

$$\varphi_j = 2\pi i \frac{1}{p_j + q_j w_j} \sim i \frac{|u_j|^2 \Im w_j}{-\Re w_j \Re u_j + \Im w_j \Im u_j + w_j \Re u_j}$$

$$= i \frac{|u_j|^2 \Im w_j}{\Im w_j \Im u_j + i \Im w_j \Re u_j} = i \frac{|u_j|^2 \Im w_j}{i \bar{u}_j \Im w_j} = u_j$$

giving $d\Phi_0 = \mathrm{id}$. $\qquad\square$

If M has a geometric decomposition $\mathcal{T}$ into ideal tetrahedra, we can easily conclude the proof of Theorem 15.1.1 as follows: the geometric ideal triangulation $\mathcal{T}$ determines a complete solution $z^0 \in \mathrm{Def}(M, \mathcal{T})$, and by Proposition 15.3.1 the image of the generalised Dehn filling map $d: \mathrm{Def}(M, \mathcal{T}) \to S^2 \times \ldots \times S^2$ is an open set U containing $(\infty, \ldots, \infty)$. The discussion in Section 14.2.7 finishes the proof.

However, we still do not know if every complete cusped finite-volume M decomposes into ideal tetrahedra! If that were the case, we would be done. Unfortunately we are not able to prove this, so more effort is needed to complete the proof.

15.3.2. Triangulations with flat tetrahedra. As we said, we are not able to prove that every cusped finite-volume M decomposes into ideal tetrahedra. The best that we can do is to show that M decomposes into ideal tetrahedra that are either ordinary or "flat", that is degenerated to an ideal quadrilateral. We show this using the Epstein–Penner canonical decomposition.

The Epstein-Penner canonical decomposition (see Theorem 5.1.21) partitions M into some ideal polyhedra $P_1, \ldots, P_k$, glued isometrically along their faces. Can we further subdivide this partition to get a geometric triangulation of M? We start by triangulating each P_i individually.

Proposition 15.3.2. *Every ideal polyhedron $P \subset \mathbb{H}^3$ subdivides into ideal tetrahedra.*

Proof. We can use the Klein model, so that ideal polyhedra in $\mathbb{H}^3$ correspond to Euclidean polyhedra with vertices in S^2. We fix a vertex $v \in P$, and a vertex v_i in every face f_i of P not incident to v. We
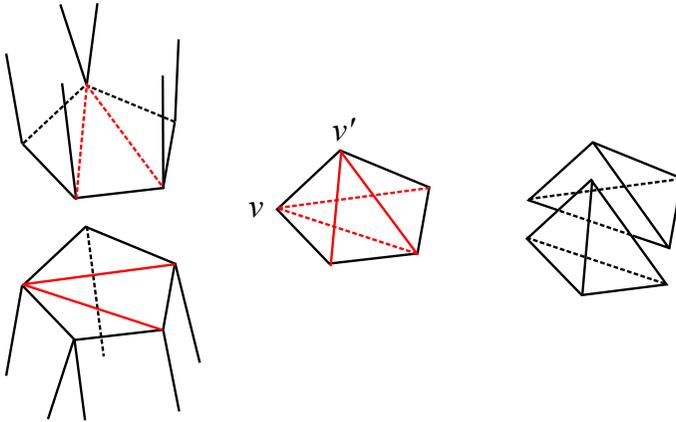
Figure 15.13. The triangulations of two isometric faces that do not match (left) because they are obtained by coning along distinct vertices $v \neq v'$ (centre). We can add some additional "flat" tetrahedra to correct this (right).

triangulate each f_i by coning along v_i, and then we triangulate P by coning all the triangulated faces f_i along v. $\square$

The problem is, that we are not able to guarantee that the triangulations of the polyhedra $P_1, \ldots, P_k$ match at each pair of glued faces. The typical situation is shown in Figure 15.13-(left): two isometric n-gons (with $n \geqslant 4$) are triangulated by coning on distinct vertices $v \neq v'$.

There is one thing that we can do, however: we can insert some *flat* tetrahedra between the two n-gons that connect the two mismatching triangulations as in Figure 15.13-(right), and obtain a partially-flat ideal triangulation of M. The flat tetrahedra can be inserted as follows: if v and v' share an edge we insert $n - 3$ flat tetrahedra as suggested by Figure 15.13-(right); otherwise, the diagonal joining v and v' is common to both triangulations and we proceed inductively on each of the two sub-polygons cut by this diagonal.

The resulting ideal triangulation $\mathcal{T}$ contains n tetrahedra $\Delta_1, \ldots, \Delta_n$; each Δ_i is either *fat* (it lies inside some polyhedron P_j) or *flat* (it connects two faces of the P_j). We now fix arbitrarily a pair of opposite edges in each Δ_i and denote by $z_i \in \mathbb{C}$ the modulus of Δ_i at these edges. If Δ_i is fat, then $\Im z_i > 0$ as usual; if Δ_i is flat, the modulus z_i still makes sense as a number in $\mathbb{R} \setminus \{0, 1\}$: the vertex triangle of Figure 14.2-(left) is a degenerated flat triangle with three vertices $0, 1, z_i$. The flat tetrahedron has three real moduli $z_i, \frac{z_i - 1}{z_i}, \frac{1}{1 - z_i}$; two arguments are 0 and one is π.

We set $z^0 = (z_1, \ldots, z_n)$. By construction, the point z^0 satisfies the consistency equations (29).

15.3.3. Perturbations. We now let U be a small contractible neighbourhood of z^0 in $(\mathbb{C} \setminus \{0, 1\})^n$. Note that U necessarily contains some points z with $\Im z_i < 0$ for some i. The consistency equations (29) make sense in U, and we define $\mathrm{Def}_U \subset U$ to be the set of solutions. We still have $\mathrm{Def}_U = G^{-1}(2\pi i, \ldots, 2\pi i)$.

Does a solution $z \in \mathrm{Def}_U$ determine a hyperbolic structure on M? Intuitively, each modulus z_j with $\Im z_j = 0$ should represent a flat tetrahedron, each z_j with $\Im z_j < 0$ a negatively oriented fat tetrahedron, and positively and negatively oriented tetrahedra should overlap; although appealing, it is however hard (and sometimes impossible) to translate this idea into a concrete definition in general.

Our aim is to prove Theorem 15.1.1 by perturbing z^0, so we only need here to furnish a geometric interpretation for the solutions z that lie in a sufficiently small neighbourhood U. We do this as follows.

If z is sufficiently close to z^0, all the fat tetrahedra that triangulate a single polyhedron P_i stay positively fat, and since they satisfy the consistency equations along the internal edges of P_i they glue to form a possibly non-convex polyhedron $P_i(z) \subset \mathbb{H}^3$. The original faces of P_i are now bent in $P_i(z)$ along their diagonals.

We now consider the originally flat tetrahedra joining two faces of two polyhedra $P_i(z)$ and $P_j(z)$. We assign all of them to one of the two polyhedra, say $P_i(z)$, and forget about $P_j(z)$. These tetrahedra are attached to some n-gonal face f of $P_i(z)$, bent along its diagonals. The effect of each attached tetrahedron Δ_j is to flip a diagonal, and we modify $P_i(z)$ accordingly by replacing the two ideal triangles in f incident to the old diagonal with the two triangles sharing the new one. Geometrically, this corresponds to adding Δ_j if $\Im z_i > 0$, or to cut Δ_j away from $P_j(z)$ if $\Im z_j < 0$. At the end of the process, we get a new polyhedron $P'_j(z)$ with the same vertices as $P_j(z)$, whose faces are bent along different diagonals.

The new possibly non-convex polyhedra $P'_j(z)$ glue isometrically along their paired faces, that are now bent along the same diagonals. The consistency equations are satisfied along these diagonals, hence we get a hyperbolic structure for M. We summarise our discoveries:

Proposition 15.3.3. *If U is sufficiently small, every solution $z \in \mathrm{Def}_U$ represents naturally a hyperbolic structure on M.*

15.3.4. Local smoothness. We have seen that every solution z that is sufficiently close to z^0 determines a hyperbolic metric on M. We now need to extend Corollary 15.2.13.

Proposition 15.3.4. *If U is sufficiently small then Def_U is a complex manifold of dimension c.*

Proof. As in Corollary 15.2.13, it suffices to prove that $dG_{z^0} : V^n \to \ker B'$ has maximum rank and is hence invertible.

The only place where we used $\Im z_j > 0$ is the proof of Proposition 15.2.12, and we now modify it so that it applies also to our partially flat solution z^0. We have $M = Z^{\mathsf{T}}(A')^{\mathsf{T}}$ and we need to prove that $\mathrm{rk}\,M = \mathrm{rk}\,A' = n - c$.

We needed $z_i \notin \mathbb{R}$ to prove that $Z^{\mathsf{T}}|_{\mathbb{R}^{2n}}$ is injective: this is not the case here, since $\frac{1}{z_i}$ and $\frac{1}{z_i-1}$ are not $\mathbb{R}$-independent when Δ_i is flat. Now we have

$$\ker Z^{\mathsf{T}}|_{\mathbb{R}^{2n}} = \mathrm{Span}\{z_i e_i + (1 - z_i)e_{i+n} \mid \Delta_i \text{ is flat}\}.$$

For $\mathbb{K} = \mathbb{R}, \mathbb{C}$ we write

$$\mathbb{K}^{2n} = \mathbb{K}^{2n}_{\text{fat}} \oplus \mathbb{K}^{2n}_{\text{flat}}, \qquad \mathbb{K}^n = \mathbb{K}^n_{\text{fat}} \oplus \mathbb{K}^n_{\text{flat}}$$

where $\mathbb{K}^{2n}_{\text{fat}}$ ($\mathbb{K}^{2n}_{\text{flat}}$) is generated by the e_i, e_{i+n} such that Δ_i is fat (flat), and similarly $\mathbb{K}^n_{\text{fat}}$ ($\mathbb{K}^n_{\text{flat}}$) is generated by the e_i such that Δ_i is fat (flat).

The following restriction of Z^{T} is still an isomorphism:

$$Z^{\mathsf{T}}|_{\mathbb{R}^{2n}_{\text{fat}}} : \mathbb{R}^{2n}_{\text{fat}} \longrightarrow \mathbb{C}^n_{\text{fat}}.$$

We use it to push-forward ω to a symplectic form Ω on $\mathbb{C}^n_{\text{fat}}$, and we extend it trivially to an alternating bilinear form Ω on $\mathbb{C}^n$ with radical $\mathbb{C}^n_{\text{flat}}$.

It is still true that $M(\mathbb{R}^{2n})$ has real dimension $n - c$, because $(A')^{\mathsf{T}}(\mathbb{R}^{2n})$ intersects $\ker Z^{\mathsf{T}}$ trivially. To prove this, suppose that a vector $v' \in \mathbb{R}^{2n}$ that is $\mathbb{R}$-generated by the rows of A' lies in the kernel of Z^{T}, that is:

- $(v'_i, v'_{i+n}) = (0, 0)$ for every fat Δ_i,
- $(v'_i, v'_{i+n}) = k_i(z_i, 1 - z_i)$ for some $k_i \in \mathbb{R}$, for every flat Δ_i.

We define $v \in \mathbb{R}^{3n}$ by setting

$$v_i = v'_i, \quad v_{n+i} = -v'_{n+i}, \quad , v_{2n+i} = 0$$

for all $i = 0, \ldots, n - 1$. We note that each triple (v_i, v_{n+i}, v_{2n+i}) is either $(0, 0, 0)$, or it contains three pairwise distinct real numbers. The vector $v' \in \mathbb{R}^{2n}$ is generated by the n rows of $A' \iff$ the vector $v \in \mathbb{R}^{3n}$ is generated by the n rows of A (exercise).

We interpret v as a colouring of the $3n$ edge pairs in $\mathcal{T}$. The colouring v is everywhere zero on all fat tetrahedra, and on each flat tetrahedron it is either $(0, 0, 0)$ or consists of three pairwise distinct numbers. We must prove that it is zero everywhere.

The vector v is contained in the lagrangian subspace generated by the rows $A_1, \ldots, A_{2n}$ of A. In particular we have $\omega(v, A_{n+i}) = 0$ for all i. Every row A_{n+i} corresponds to an edge e of $\mathcal{T}$, and this equality says that the colours on all tetrahedra incident to e of the edge pairs not containing e sum to zero, provided that they are counted with signs in each tetrahedron. This implies the following: if all such tetrahedra except one Δ_j have colour zero on all their edge pairs, then two edge pairs on Δ_j have the same colours, and by what just said on v this implies that all colours are zero also on Δ_j.

Consider a sequence $\Delta_1, \ldots, \Delta_k$ of flat tetrahedra that connects two faces of two polyhedra P_i and P_j of the Epstein-Penner decomposition. There is one edge of Δ_1 that is incident only to fat tetrahedra except Δ_1, and the above discussion implies that all the colours on Δ_1 are zero; by induction we deduce that all the flat Δ_i have zero colours. Therefore $v = 0$, as required.

We have proved that $\dim M(\mathbb{R}^{2n}) = n - c$, and we conclude as in the proof of Proposition 15.2.12. The form Ω defines a positive semidefinite scalar product $\langle , \rangle$ on $\mathbb{C}^n$ with radical $\mathbb{C}^n_{\text{flat}}$. If $\mathrm{rk} M < n - c$, there are two vectors $v_1, v_2 \in \mathbb{R}^n$ such that $M(v_1 + iv_2) = 0$ while $Mv_1 \neq 0$ and $Mv_2 \neq 0$. We get $\|Mv_1\| = \|Mv_2\| = 0$ as in the proof of Proposition 15.2.12, and this implies that Mv_1, Mv_2 both lie in the radical $\mathbb{C}^n_{\text{flat}}$.

If $Mv_1, Mv_2 \in \mathbb{C}^n_{\text{flat}}$ then both Mv_1 and Mv_2 are real since $\frac{1}{z_i}, \frac{1}{z_i - 1} \in \mathbb{R}$ for every flat Δ_i. Now $0 = M(v_1 + iv_2) = Mv_1 + iMv_2$ implies that $Mv_1 = Mv_2 = 0$, a contradiction. $\qquad\square$

15.3.5. Conclusion of the proof. The proof of Theorem 15.1.1 finishes like in the geometric triangulation case. All the discussion of Section 15.2.7 applies here: the holonomies are defined also in this context and may be used to parametrize the complex manifold $U(z^0)$. The strategy of Section 15.3.1 is still valid: the generalised Dehn filling invariant map d sends $U(z^0)$ to an open neighbourhood of $(\infty, \ldots, \infty)$ in $S^2 \times \ldots \times S^2$ and the discussion in Section 14.2.7 finishes the proof. We have completed the proof of Theorem 15.1.1.

15.3.6. Mostow–Prasad rigidity. We can use the Dehn filling Theorem 15.1.1 to extend Mostow Rigidity Theorem 13.3.1 from the closed to the cusped case. The cusped case was proved by Prasad in 1973, hence the theorem is known as *Mostow–Prasad Rigidity Theorem*.

Let N a be compact orientable three-manifold bounded by some $c \geqslant 0$ tori, and let $M = \mathrm{int}(N)$.

Theorem 15.3.5 (Mostow–Prasad rigidity). *The manifold M admits at most one finite-volume complete hyperbolic metric up to isometries homotopic to the identity.*

Proof. The $c = 0$ case is Theorem 13.3.1, so we suppose $c \geqslant 1$. Roughly, the metric on M is unique because it is the limit of hyperbolic metrics on its closed Dehn fillings, that are unique by Theorem 13.3.1. (We are using the hyperbolic Dehn filling theorem as a trick to extend Mostow's rigidity from the closed to the cusped case: this is certainly not Prasad's original proof, that was published before any hyperbolic Dehn filling was studied, and more importantly it applies to all dimensions $n \geqslant 3$.)

More precisely, let M have two finite-volume hyperbolic metrics m_1 and m_2. The Dehn filling Theorem holds for both, and yields sequences M_i^j of closed hyperbolic three-manifolds obtained from M by Dehn filling with parameter s^j such that $s^j \to (\infty, \ldots, \infty)$, for $i = 1, 2$.

Mostow theorem furnishes an isometry $M_1^j \to M_2^j$ homotopic to the identity. The core closed geodesics of the filling solid tori are unique in their homotopy classes, hence by removing them we get isometric incomplete metrics. Since m_1 and m_2 are limits of these, they are also isometric.

This latter sentence can be made rigorous as follows: let $\mathcal{T}$ be a partially flat ideal triangulation for M in the metric m_1. Pick an ideal tetrahedron Δ in $\mathcal{T}$ and note that it can be straightened uniquely also in the incomplete metrics for M close to the complete ones, just by extending the construction in the proof of Proposition 7.4.9 from dimension two to three (an orientation of the core geodesics is needed to define the straightening unambiguously). The modulus $z \in \mathbb{C}_{>0} \cup (\mathbb{R} \setminus \{0, 1\})$ of the straightened Δ in M is intrinsically determined as the limit of the moduli of the straightenings of Δ in the M_i^j, hence it is the same for both metrics m_1 and m_2. It follows that m_2 has the same geometric ideal decomposition of m_1, and hence $m_1 = m_2$ after a homotopy. $\square$

15.4. Volumes

We have discovered that "most" Dehn fillings of a cusped finite-volume hyperbolic manifold are hyperbolic, and we now discuss their volumes.

Let M be a complete orientable finite-volume cusped hyperbolic three-manifold. We say that a Dehn filling of M is *non-trivial* if it fills at least one boundary torus; in other words, it is determined by some Dehn filling parameter s different from $(\infty, \ldots, \infty)$, after fixing some arbitrary homology bases for the boundary tori. We prove in this section the following theorem.

Theorem 15.4.1. *If a non-trivial Dehn filling M^{fill} of M admits a complete finite-volume hyperbolic metric, then*

$$\mathrm{Vol}(M^{\mathrm{fill}}) < \mathrm{Vol}(M).$$

If s^i is any sequence of non-trivial Dehn filling parameters converging to $(\infty, \ldots, \infty)$, and M^i is the manifold obtained by filling M along s^i, then M^i is eventually hyperbolic and

$$\mathrm{Vol}(M^i) \nearrow \mathrm{Vol}(M).$$

The cores of the filling solid tori are eventually simple closed geodesics and their lengths tend to zero.

We already know that if $s^i \to (\infty, \ldots, \infty)$ then $\mathrm{Vol}(M^i) \to \mathrm{Vol}(M)$ and the cores of the solid tori are closed geodesics whose length tend to zero: this follows from the construction of the metric on M^i via geometric ideal triangulations. In some sense (that we will not make precise), the metric on M^i converges to that of M, and the core geodesics become so short that they disappear in the limit producing cusps.

The only new fact to prove is that the volume strictly decreases under *any* Dehn filling M^{fill}, not only those that are close to $(\infty, \ldots, \infty)$: this is

rather subtle, because the hyperbolic metric on M^{fill} might not be obtained by completing ideal triangulations of M. (For instance, the core tori in M^{fill} may not be isotopic to closed geodesics.)

We prove the theorem as follows. We first consider the case where M has an ideal geometric triangulation $\mathcal{T}$: in that case $\text{Def}(M, \mathcal{T})$ is a nice smooth complex manifold and we investigate the volume function

$$\text{Vol}: \text{Def}(M, \mathcal{T}) \longrightarrow \mathbb{R}_{>0}$$

that computes the volume of each hyperbolic structure on M. We show that the volume function makes sense in a bigger convex set containing $\text{Def}(M, \mathcal{T})$ and is also easier to study there. This bigger convex set is the set of *angle structures* on $\mathcal{T}$. We prove that the volume function is strictly concave there, and the complete solution is a global maximum.

We conclude with a couple of finite-covering tricks to pass from local (near $(\infty, \ldots, \infty)$) to global, and to deal with more general manifolds M that do not decompose into hyperbolic ideal tetrahedra.

15.4.1. Angle structures. Let $M = \text{int}(N)$ where N is an orientable compact three-manifold bounded by $c > 0$ tori, equipped with an oriented ideal triangulation $\mathcal{T}$ with n tetrahedra $\Delta_1, \ldots, \Delta_n$ and n edges $e_1, \ldots, e_n$. We identify orientation-preservingly each Δ_i with the one shown in Figure 15.5, and we label the edge pairs of type 1, 2, 3, with the numbers i, $i+n$, and $i+2n$ respectively.

An *angle structure* θ on M is the assignment of an *angle* $0 < \theta_j < \pi$ to the j-th edge pair in $\mathcal{T}$, with the following requirements:

(1) at each tetrahedron Δ_i we have $\theta_i + \theta_{i+n} + \theta_{i+2n} = \pi$,
(2) at each edge e_i, the sum of the incident angles is 2π.

The set of angle structures is thus a subset $\mathcal{A} \subset \mathbb{R}_{>0}^{3n}$. It is the intersection of the affine subspace in $\mathbb{R}^{3n}$ determined by the conditions (1) and (2) with the cone $\mathbb{R}_{>0}^{3n}$, so in particular it is convex. By definition this affine subspace is the solution space of the linear system

$$A\theta = \begin{pmatrix} \pi \\ 2\pi \end{pmatrix}$$

where A is the incidence $2n \times 3n$ defined in Section 15.2.2 and depicted in Figure 15.6-(left), and $\theta = (\theta_1, \ldots, \theta_{3n}) \in \mathbb{R}^{3n}$.

What is the geometric meaning of an angle structure θ on $\mathcal{T}$? Thanks to condition (1), every Δ_i is realised by a unique ideal hyperbolic tetrahedron with dihedral angles θ_i, θ_{i+n}, and θ_{i+2n} whose complex modulus is

$$(31) \qquad z_i = \frac{\sin \theta_{i+n}}{\sin \theta_{i+2n}} e^{i\theta_i}.$$

Angle structures are then in natural 1-1 correspondence with the realisations of the tetrahedra $\Delta_1, \ldots, \Delta_n$ as ideal hyperbolic tetrahedra, with the mild condition (2) that their dihedral angles around edges must sum to 2π.

Condition (2) is only "half" of the consistency equation at each e_i, which requires the product of the complex moduli to be 1: condition (2) alone does *not* guarantee that by gluing these tetrahedra we get a hyperbolic structure on M: some "shearing" may arise around an edges.

With this geometric description, the deformation space $\mathrm{Def}(M, \mathcal{T})$ is naturally a smooth submanifold of $\mathcal{A}$.

15.4.2. Dimensions. We will henceforth suppose that $\mathrm{Def}(M, \mathcal{T})$ is non-empty and contains a complete solution z^0, that is $\mathcal{T}$ can be realised as a geometric ideal triangulation. This implies in particular that $\mathcal{A}$ is also non-empty and has dimension $\dim \ker A = 3n - (2n - c) = n + c$ by Corollary 15.2.8. The space $\mathrm{Def}(M, \mathcal{T})$ is a real smooth submanifold of dimension $2c$.

The tangent space $T_\theta \mathcal{A}$ at any point $\theta \in \mathcal{A}$ is of course $\ker A$. We now construct an explicit set of generators for $\ker A$. Recall from Section 15.2.6 that every simplicial closed oriented curve γ in a triangulated boundary torus T_i defines an integer vector $v_\gamma \in \mathbb{R}^{3n}$.

Recall the alternating form ω on $\mathbb{R}^{3n}$ from Section 15.2.3: we consider it as a matrix and write $v^* = \omega v$ for all $v \in \mathbb{R}^{3n}$, so that $\omega(v, w) = \langle v, w^* \rangle$ where $\langle , \rangle$ is the Euclidean scalar product. We denote by A_i the i-th row of A, considered as a vector in $\mathbb{R}^{3n}$.

We fix two arbitrary generators m_i, l_i of $\pi_1(T_i)$ for all $i = 1, \ldots, c$.

Proposition 15.4.2. *The tangent space $T_\theta \mathcal{A} = \ker A$ is generated by the (not independent) $n + 2c$ vectors $A^*_{n+1}, \ldots, A^*_{2n}, v^*_{m_1}, v^*_{l_1}, \ldots, v^*_{m_c}, v^*_{l_c}$.*

Proof. Propositions 15.2.9 and 15.2.14 say that the rows of A are ω-orthogonal to themselves and to v_γ for any γ, hence they are $\langle , \rangle$-orthogonal to all vectors A^*_i and v^*_γ, in other words $A^*_i, v^*_\gamma \in \ker A$.

To prove that the $n + 2c$ listed vectors generate $\ker A$ it suffices to show that they span a subspace of dimension $n + c$. Proposition 15.2.15 says that if we add the rows $v_{m_1}, \ldots, v_{m_c}$ to A we get a rank-$2n$ matrix. The same proof shows that by further adding the rows $v_{l_1}, \ldots, v_{l_c}$ we get a rank-$(2n + c)$ matrix.

The vectors $A_1, \ldots, A_{2n}, v_{m_1}, v_{l_1}, \ldots, v_{m_c}, v_{l_c}$ span a dimension-$(2n + c)$ space. Since $\dim \ker \omega = n$, the vectors $A^*_1, \ldots, A^*_{2n}, v^*_{m_1}, v^*_{l_1}, \ldots, v^*_{m_c}, v^*_{l_c}$ span a space of dimension at least $2n + c - n = n + c$. Since $A^*_1 = \ldots = A^*_n = 0$, we are done. $\qquad\square$

15.4.3. The volume function. The volume function on $\mathrm{Def}(M, \mathcal{T})$ extends naturally to a function

$$\mathrm{Vol} \colon \mathcal{A} \longrightarrow \mathbb{R}_{>0}$$

which assigns to each angle structure θ the sum of the volumes of the ideal hyperbolic realisations of $\Delta_1, \ldots, \Delta_n$ determined by θ. Recall from Section 13.1 that the volume of Δ_i is

$$\mathrm{Vol}(\Delta_i) = \Lambda(\theta_i) + \Lambda(\theta_{i+n}) + \Lambda(\theta_{i+2n})$$

where Λ is the Lobachevsky function

$$\Lambda(\alpha) = -\int_0^\alpha \log|2\sin t|\,dt.$$

We therefore get a simple-looking formula

$$\mathrm{Vol}(\theta) = \sum_{j=1}^{3n} \Lambda(\theta_j)$$

which implies immediately the following important fact.

Proposition 15.4.3. *For every $\theta \in \mathcal{A}$ and $v \in T_\theta\mathcal{A}$ we have*

$$\frac{\partial \mathrm{Vol}}{\partial v} = \sum_{j=1}^{3n} -v_j \log\sin\theta_j, \qquad \frac{\partial^2\mathrm{Vol}}{\partial v^2} < 0.$$

Proof. We consider a single addendum $\Lambda(\theta_j)$ of $\mathrm{Vol}(\theta)$. We have

$$\frac{\partial\Lambda(\theta_j)}{\partial v} = \Lambda'(\theta_j)\frac{\partial\theta_j}{\partial v} = -\log|2\sin\theta_j|v_j = -v_j\log 2 - v_j\log\sin\theta_j.$$

Since $v \in T_\theta\mathcal{A}$ we have $v_i + v_{i+n} + v_{i+2n} = 0$ for all i and hence $\sum_j v_j \sin 2 = 0$. This proves the first equality. To estimate the second derivative, we may suppose up to symmetries that $\theta_i, \theta_{i+n} < \frac{\pi}{2}$ and we get

$$-\frac{\partial^2\mathrm{Vol}(\Delta_i)}{\partial v^2} = v_i^2 \cot\theta_i + v_{i+n}^2 \cot\theta_{i+n} + v_{i+2n}^2 \cot\theta_{i+2n}$$

$$= v_i^2\cot\theta_i + v_{i+n}^2\cot\theta_{i+n} + (v_i + v_{i+n})^2\frac{1-\cot\theta_i\cot\theta_{i+n}}{\cot\theta_i+\cot\theta_{i+n}}$$

$$= \frac{(v_i+v_{i+n})^2 + (v_i\cot\theta_i - v_{i+n}\cot\theta_{i+n})^2}{\cot\theta_i+\cot\theta_{i+n}} > 0.$$

Indeed the denominator is positive since $\theta_i, \theta_{i+n} < \frac{\pi}{2}$, and if the numerator is 0 then $v_i = -v_{i+n}$ and hence $\cot\theta_i = -\cot\theta_{i+n}$, a contradiction. $\qquad\square$

We have proved that Vol is a smooth strictly concave function on $\mathcal{A}$.

15.4.4. The complete solution. We now remember that $\mathrm{Def}(M, \mathcal{T})$ contains a complete solution z^0 by assumption.

Proposition 15.4.4. *The function $\mathrm{Vol}\colon \mathcal{A} \to \mathbb{R}_{>0}$ has a unique global maximum at the complete solution z^0.*

Proof. The complete solution z^0 has an angle structure θ. Proposition 15.4.3 says that

$$\frac{\partial\mathrm{Vol}}{\partial v} = \sum_{j=1}^{3n} -v_j\log\sin\theta_j$$

for all $v \in T_\theta\mathcal{A}$. At every edge e_i, the consistency equation is of type $\sum_j \log w_j = 2\pi i$ where $w_1, \ldots, w_k$ are the moduli of the incident tetrahedra. The imaginary part of this equation is satisfied at every point of $\mathcal{A}$,

so we look at the real part and using (31) we discover that

$$0 = \Re \sum_j \log w_j = \sum_j \log |w_j| = \sum_j \left(\log \sin \alpha_j^1 - \log \sin \alpha_j^2 \right)$$

where α_j^1 and α_j^2 are the other two angles in the tetrahedron with modulus w_j, in counterclockwise order. The latter expression is equivalent to

$$-\sum_{j=1}^{3n} (A_{n+i}^*)_j \log \sin \theta_j = \frac{\partial \mathrm{Vol}}{\partial A_{n+i}^*}$$

which is hence zero for all $i = 1, \ldots, n$.

We make some analogous considerations for the completeness equations. Since z^0 is complete, for every simplicial boundary curve γ we have $\sum_j \log w_j = |\gamma|\pi i$ where $w_1, \ldots, w_k$ are the moduli encountered by γ at its right, and the real part of this equation implies that $\frac{\partial \mathrm{Vol}}{\partial v_\gamma^*} = 0$.

Proposition 15.4.2 shows that vectors of type A_{n+i}^* and v_γ^* generate $T_\theta \mathcal{A}$, hence the gradient of Vol vanishes at θ. Since Vol is strictly concave, this critical point is the unique global maximum. $\qquad\square$

Corollary 15.4.5. *The space* $\mathrm{Def}(M, \mathcal{T})$ *contains at most one complete solution.*

15.4.5. Volumes and Dehn filling. We can now prove Theorem 15.4.1 on manifolds having a geometric ideal triangulation. We start with a "local" version which also considers generalised Dehn fillings, see Section 15.1.1.

Lemma 15.4.6. *If M has a geometric ideal triangulation, there is a neighbourhood U of $(\infty, \ldots, \infty)$ such that* $\mathrm{Vol}(M^{\mathrm{fill}}) < \mathrm{Vol}(M)$ *for every non-trivial generalised Dehn filling parameter $s \in U$. Moreover* $\mathrm{Vol}(M^{\mathrm{fill}}) \nearrow \mathrm{Vol}(M)$ *as s tends to* $(\infty, \ldots, \infty)$.

Proof. The Dehn Filling Theorem 15.1.1 says that all Dehn fillings in an open neighbourhood U of $(\infty, \ldots, \infty)$ are hyperbolic and constructed by completing solutions in $\mathrm{Def}(M, \mathcal{T})$, and they have smaller volume than M in virtue of Proposition 15.4.4. (The completion adds a circle to M which does not contribute to the volume.)

The volume of M^{fill} depends smoothly on $s \in U$ and has a global maximum at $(\infty, \ldots, \infty)$. In particular, when $s \to (\infty, \ldots, \infty)$ we get $\mathrm{Vol}(M^{\mathrm{fill}}) \to \mathrm{Vol}(M)$ from below. $\qquad\square$

We now need to drop the open set U and prove that volume decreases under *any* Dehn filling of M.

Lemma 15.4.7. *Let M have a geometric ideal triangulation. If a nontrivial Dehn filling M^{fill} admits a complete finite-volume hyperbolic metric, then* $\mathrm{Vol}(M^{\mathrm{fill}}) < \mathrm{Vol}(M)$.

Proof. Let $s = (s_1, \ldots, s_c)$ be the Dehn filling parameter giving M^{fill} and let U be the open neighbourhood of $(\infty, \ldots, \infty)$ furnished by Lemma 15.4.6. For sufficiently big integer $k > 0$ we have $ks \in U$ and hence M^{fill} also admits a hyperbolic metric with cone angles $\frac{2\pi}{k}$ on the core geodesics of the filling solid tori, with volume strictly smaller than $\text{Vol}(M)$. We may interpret this metric as a hyperbolic orbifold O, and we have $\text{Vol}(O) < \text{Vol}(M)$.

If M^{fill} is closed Proposition 13.3.8 gives

$$\text{Vol}(M) > \text{Vol}(O) \geqslant v_3 \|M^{\text{fill}}\| = \text{Vol}(M^{\text{fill}})$$

and we are done.

If M^{fill} is not closed, we can find some closed fillings of M^{fill} with volume arbitrarily close to $\text{Vol}(O)$ that can be interpreted as orbifolds, and get $\text{Vol}(O) \geqslant \text{Vol}(M^{\text{fill}})$ as a limit. $\qquad\square$

The proof of Lemma 15.4.7 shows also the following.

Lemma 15.4.8. *Let M have a geometric ideal triangulation. We have* $v_3 \|M^{\text{fill}}\| < \text{Vol}(M)$ *for every closed Dehn filling M^{fill} of M.*

We prove a lemma that will be useful below.

Lemma 15.4.9. *Let M have a geometric ideal triangulation. If M_i is a hyperbolic Dehn filling of M with parameter s^i and $\text{Vol}(M_i) \to \text{Vol}(M)$, then $s^i \to (\infty, \ldots, \infty)$.*

Proof. We may suppose that the M_i are closed, since every sequence of Dehn fillings can be approximated by a sequence of closed ones. If s^i lies definitely in the open set U furnished by Theorem 15.1.1, the assertion follows because the volume function there has $\text{Vol}(M)$ as a unique maximum (thanks to Proposition 15.4.4).

Otherwise, there is a fixed big $k \in \mathbb{N}$ such that $ks^i \in U$ for all i on a subsequence, but ks^i does not converge to $(\infty, \ldots, \infty)$. Thus M_i admits a hyperbolic orbifold structure O_i with $\text{Vol}(M) - \varepsilon > \text{Vol}(O_i)$ and Proposition 13.3.8 gives $\text{Vol}(M) - \varepsilon > \text{Vol}(O) \geqslant v_3 \|M_i\| = \text{Vol}(M_i)$ $\forall i$. $\qquad\square$

Again, the proof shows also the following.

Lemma 15.4.10. *Let M have a geometric ideal triangulation. If M_i is a closed Dehn filling of M with parameter s^i and $v_3 \|M_i\| \to \text{Vol}(M)$, then $s^i \to (\infty, \ldots, \infty)$.*

To conclude the proof of Theorem 15.4.1 we only need to consider the unlucky case where M has no geodesic ideal triangulations. We prove that geometric triangulations exist on some finite cover of M, and this will suffice for our purposes.

15.4.6. Geometric triangulations exist virtually. We prove here the following result.

Theorem 15.4.11. *Every cusped complete finite-volume 3-manifold M has a finite-sheeted cover $\tilde{M}$ with a geometric ideal triangulation $\mathcal{T}$.*

Proof. The Epstein-Penner decomposition subdivides M into ideal polyhedra $P_1, \ldots, P_k$. If distinct vertices of P_i lie in distinct cusps of M for all i, then the decomposition can be easily subdivided into an ideal triangulation: we order the cusps, so the vertices of each P_i inherit an ordering which is preserved along the matching faces; we subdivide each P_i as in the proof of Proposition 15.3.2 by coning on the smallest vertex at each step: the triangulations on the paired faces match and we get a geometric ideal triangulation for M itself.

We now show that some cover $\tilde{M}$ has a decomposition of this type. The decomposition of M into $P_1, \ldots, P_k$ lifts in any degree-d cover $\tilde{M}$ to a decomposition of $\tilde{M}$ into $P_1^i, \ldots, P_k^i$ with $i = 1, \ldots, d$. We now prove that there is a $\tilde{M}$ such that distinct vertices of P_j^i lie in distinct cusps, for all i, j.

To do so, we prove that for every pair of vertices $v \neq v'$ of some P_i there is a covering where the lifts $\tilde{v}, \tilde{v}'$ in P_i^j for some j lie in distinct cusps. This is enough to conclude: pick a *regular* covering $\tilde{M}$ that covers all these finitely many coverings; since its desk transformation group acts transitively on lifts, all distinct vertices on every P_i^j lie in distinct cusps.

We have $M = \mathbb{H}^3 / \Gamma$. We fix an arbitrary lift $\tilde{P}_i \subset \mathbb{H}^3$ of P_i. Two vertices $v \neq v'$ of P_i lie in the same cusp $T \Leftrightarrow$ there is a $\gamma \in \Gamma$ such that $\gamma(\tilde{v}) = \tilde{v}'$, where $\tilde{v}, \tilde{v}' \in \tilde{P}_i$ are the lifts of v, v'. The subgroup $\mathrm{Stab}_\Gamma(\tilde{v}) = \pi_1(T)$ is separable by Corollary 4.3.13 and hence there is a finite-index $H < \Gamma$ that contains $\mathrm{Stab}_\Gamma(\tilde{v})$ but avoids γ.

No element $\varphi \in H$ is such that $\varphi(\tilde{v}) = \tilde{v}'$, otherwise we would have $\gamma^{-1}\varphi \in \mathrm{Stab}_\Gamma(v) < H$ and hence $\gamma \in H$, a contradiction. We have constructed a finite-cover $\mathbb{H}^3 / H$ where in some lift P_i^j of P_i the two lifted vertices $\tilde{v}, \tilde{v}'$ lie in distinct cusps. $\square$

15.4.7. Conclusion of the proof. We can now prove Theorem 15.4.1.

Proof of 15.4.1. Some degree-d cover $\tilde{M}$ of M has an ideal geometric triangulation, hence the theorem holds for $\tilde{M}$.

For every Dehn filling M^{fill} of M, there is a Dehn filling $\tilde{M}^{\mathrm{fill}}$ of $\tilde{M}$ such that the cover $\tilde{M} \to M$ extends to a degree-d map $\tilde{M}^{\mathrm{fill}} \to M^{\mathrm{fill}}$.

If M^{fill} is closed hyperbolic, we have

$$d\mathrm{Vol}(M) = \mathrm{Vol}(\tilde{M}) > v_3 \|(\tilde{M}^{\mathrm{fill}})\| \geqslant d v_3 \|M^{\mathrm{fill}}\| = d\mathrm{Vol}(M^{\mathrm{fill}})$$

using Lemma 15.4.8, hence $\mathrm{Vol}(M) > \mathrm{Vol}(M^{\mathrm{fill}})$.

If M^{fill} is cusped hyperbolic, it can be approximated by closed hyperbolic fillings M_i^{fill}, and there are closed fillings $\tilde{M}_i^{\mathrm{fill}}$ of $\tilde{M}^{\mathrm{fill}}$ with degree-d

maps $\tilde{M}_i^{\text{fill}} \to M_i^{\text{fill}}$. By Lemma 15.4.10 there is an $\varepsilon > 0$ such that

$$\text{Vol}(\tilde{M}) - \varepsilon > v_3 \|\tilde{M}_i^{\text{fill}}\| \geqslant dv_3 \|M_i^{\text{fill}}\| = d\text{Vol}(M_i^{\text{fill}})$$

for all i, hence $d\text{Vol}(M) - \varepsilon = \text{Vol}(\tilde{M}) - \varepsilon \geqslant d\text{Vol}(M^{\text{fill}})$. $\qquad\qquad$ $\square$

15.4.8. Bounded volume. How many hyperbolic manifolds are there with bounded volume? In general, infinitely many.

We have just discovered that there are infinitely many hyperbolic manifolds with volume smaller than $2v_3 = 2.0298832128\ldots$ because the figure-eight knot complement M has $\text{Vol}(M) = 2v_3$ and every hyperbolic Dehn filling of M has volume smaller than $\text{Vol}(M)$. Moreover, the volumes of the Dehn fillings of M form an indiscrete set that tends to $2v_3$ from below.

Theorem 15.4.1 shows that this is the typical situation: the hyperbolic Dehn filling produces infinite sequences of manifolds with bounded volume. Despite this variety of manifolds, we can still control topologically the hyperbolic manifolds having bounded volume, at least in principle.

Theorem 15.4.12. *For every $V > 0$ there is a compact 3-manifold N bounded by tori, such that every complete finite-volume orientable hyperbolic 3-manifold M with $\text{Vol}(M) < V$ is diffeomorphic to the interior of some Dehn filling of N.*

Proof. This is a consequence of the thick-thin decomposition. Let $\varepsilon > 0$ be a fixed Margulis constant: every complete finite-volume M decomposes along disjoint embedded tori as $M = M^{\text{thick}} \cup M^{\text{thin}}$ where M^{thick} is compact and has injectivity radius $> \varepsilon$, and M^{thin} consists of truncated cusps and tube neighbourhoods of short geodesics. In particular M is a Dehn filling of M^{thick}.

We now show that there are only finitely many possible diffeomorphism types for M^{thick} with volume $< V$. This concludes the proof: by Proposition 11.3.18 every such diffeomorphism type is realised from a link $L \cup C \subset S^3$ by surgerying along L and drilling along C; we take a disjoint union of all these links in S^3, and the complement of the resulting link in S^3 is our N.

Let $X \subset M^{\text{thick}}$ be a *maximal $\frac{\varepsilon}{2}$-net*, that is a finite set of points that stay at pairwise distance $\geqslant \frac{\varepsilon}{2}$, such that every other point in M^{thick} stays at distance $< \frac{\varepsilon}{2}$ from X (X exists because M^{thick} is compact). The $\frac{\varepsilon}{4}$-balls centred at the points in X are embedded and disjoint: hence X has cardinality at most $C(V) = V/\text{Vol}(B(x, \frac{\varepsilon}{4}))$.

The set X determines a cellularisation of M^{thick} as follows. Lift X to $\tilde{X} \subset \mathbb{H}^3$, take the Voronoi tessellation (see Section 3.3.4) determined by $\tilde{X}$ and project it back to M: this furnishes a subdivision of M^{thick} into polyhedra. Since X is maximal, every polyhedron has at most $C'(\varepsilon)$ faces with C' depending only on ε. Hence there are finitely many combinatorial types of polyhedra, depending only on ε. With at most $C(V)$ of them we

obtain only finitely many manifolds. (The proof of Proposition 11.1.6 applies to polyhedra and shows that the combinatorial subdivision determines the smooth manifold.)   □

15.4.9. Volumes of hyperbolic three-manifolds. We finish this chapter by furnishing some qualitative information on the set of volumes of hyperbolic three-manifolds. We start with the following.

Proposition 15.4.13. *Let M_i be a sequence of non-diffeomorphic complete orientable hyperbolic three-manifolds of uniformly bounded volume. After passing to a subsequence, we may suppose that there is a hyperbolic manifold M such that each M_i is a Dehn filling of M with parameter $s^i \to (\infty, \ldots, \infty)$.*

Proof using geometrisation. We know from Theorem 15.4.12 that there is a manifold N such that each M_i is a Dehn filling of N with parameter s^i. After passing to a subsequence, we may suppose that at each boundary torus T_j of N the slope s_j^i is either constant or goes to infinity. Let M be obtained from N by Dehn filling the slopes that stay constant.

If M is hyperbolic, we are done. Unfortunately, this may not be the case, and we use geometrisation to solve this annoying possibility (the original proof of Jørgensen and Thurston [**56**] does not need geometrisation and follows from a discussion on the geometric convergence of hyperbolic manifolds: we employ geometrisation only for simplicity). If M is not prime, it contains some essential spheres that always bound balls in the filled M_i and hence we may cut M along them and cup off with balls (reducing the components of ∂M). Now M is prime and its geometric decomposition must contain some hyperbolic block, otherwise M would be a graph manifold and no Dehn filling of M would be hyperbolic (it would be a graph manifold again). The tori of the decomposition of M become compressible in M_i, so up to passing to a subsequence we may suppose that each M_i is a Dehn filling of one fixed hyperbolic block of M. If s_i does not tend to $(\infty, \ldots, \infty)$ in this new setting, we restart from the beginning and proceed by induction on the number of boundary components of M.   □

Corollary 15.4.14. *The set of volumes of all complete orientable hyperbolic three-manifolds is well-ordered. For every value V there are only finitely many such manifolds with volume V.*

Proof. Suppose that there are M_i with $\mathrm{Vol}(M_1) > \mathrm{Vol}(M_2) > \ldots$ By the previous proposition these are all obtained by Dehn filling the same hyperbolic M with $s^i \to (\infty, \ldots, \infty)$, and Theorem 15.4.1 says they should converge to $\mathrm{Vol}(M)$ from below, a contradiction. The same argument show that there are finitely many manifolds of any fixed volume.   □

We deduce that the volumes of all the complete orientable hyperbolic three-manifolds are indexed by countable ordinals. The volumes $v_1 < v_2 <$
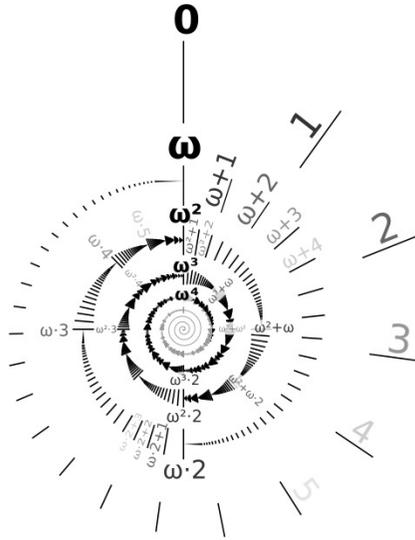
Figure 15.14. The volumes of complete orientable hyperbolic 3-manifolds form a well-ordered subset of $\mathbb{R}$ of type ω^ω.

..., of the smallest closed manifolds converge to the volume v_ω of the smallest 1-cusped manifold; the volumes $v_\omega < v_{2\omega} < \ldots$ of the smallest 1-cusped manifolds converge to the volume v_{ω^2} of the smallest 2-cusped manifold, and so on. The set of volumes has order type ω^ω, see Figure 15.14. We can also deduce the following.

Corollary 15.4.15. *For every $V > 0$ and $R > 0$ there are finitely many closed hyperbolic 3-manifolds M with $\mathrm{Vol}(M) < V$ and $\mathrm{inj}(M) > R$.*

Proof. On a sequence of closed hyperbolic 3-manifolds with bounded volume, the length of some core geodesic tends to zero. □

15.4.10. References. To furnish a complete proof of Thurston's Dehn filling Theorem we have patched together a number of different sources. The most relevant ones are as usual Thurston's notes [**56**], Benedetti – Petronio [**4**], and Ratcliffe [**49**], together with Neumann – Zagier [**42**].

Some arguments have been updated here in light of some new results that were discovered more recently by some authors. The list of non-exceptional slopes for the figure-eight knot and the Whitehead link can be found in Martelli – Petronio [**38**]. The smoothness of the deformation space stated in Corollary 15.2.13 has been proved by Choi [**13**] in 2004. Many of the arguments of this chapter, including all the discussion on angle structures, have been taken from a nice paper of Futer – Guéritaud [**21**].

The first complete proof of Thurson's Dehn filling Theorem, that takes care of the annoying case of triangulations with flat tetrahedra, is due to Petronio – Porti [**48**]. We diverge from that proof in the last arguments, where we extend Choi's smoothness theorem to this partially flat context in Proposition 15.3.4 (this proposition is probably the only original result of this book). The virtual existence of geometric triangulations was proved by Luo – Schleimer – Tillmann [**37**] in 2008.

Bibliography

[1] I. Agol, *The minimal volume orientable hyperbolic 2-cusped 3-manifolds*, Proc. Amer. Math. Soc. **138** (2010), 3723–3732.

[2] I. R. Aitchison − J. H. Rubinstein, *Localising Dehn's lemma and the loop theorem in 3-manifolds*, Math. Proc. Cambrdge Phil. Soc. **137** (2004), 281 − 292.

[3] J. Aramayona − C. J. Leininger, *Hyperbolic structures on surfaces and geodesic currents*, in "Algorithmic and Geometric Topics Around Free Groups and Automorphisms", with V. Diekert, C. Leininger, P. Silva, A. Weiss. Advanced Courses in Math (CRM), Birkhauser.

[4] R. Benedetti − C. Petronio, "Lectures on hyperbolic geometry," Universitext, Springer-Verlag, 1991.

[5] M. Bestvina, K. Bromberg, K. Fujiwara, J. Souto, *Shearing coordinates and convexity of length functions on Teichmüller space*, American Journal of Math. **135** (2013), 1449–1476.

[6] F. Bonahon. *Bouts des variétés hyperboliques de dimension 3*, Ann. of Math. **124** (1986), 71–158.

[7] ———, *The geometry of Teichmüller space via geodesic currents*, Invent. Math. **92** (1988), 139–162.

[8] ———, *Geometric structures on 3-manifolds*, Handbook of Geometric Topology (2002), Elsevier, 93–164.

[9] D. Calegari, "Foliations and the Geometry of 3-Manifolds," Oxford University Press, 2007.

[10] P. J. Callahan − M. V. Hildebrand − J. R. Weeks, *A census of cusped hyperbolic 3-manifolds*, Math. Comp. **68** (1999), 321-332.

[11] C. Cao − G. R. Meyerhoff, *The orientable cusped hyperbolic 3-manifolds of minimum volume*, Invent. Math. **146** (2001), 451–478.

[12] A. J. Casson − S. A. Bleiler, "Automorphisms of surfaces after Nielsen and Thurston", London Mathematical Society Student Texts **9**, Cambridge University Press, Cambridge, 1988.

[13] Y-E. Choi, *Positively oriented ideal triangulations on hyperbolic three-manifolds*, Topology **43** (2004), 1345–1371.

[14] H. Coxeter, "Regular Polytopes," Dover, II edition, 1963.

[15] M. Culler − N. Dunfield − J. Weeks, SnapPy, a computer program for studying the geometry and topology of 3-manifolds, `http://www.math.uic.edu/t3m/SnapPy/`

[16] M. P. Do Carmo, "Riemannian Geometry," Birkhäuser, 1992.

[17] D. B. A. Epstein − R. C. Penner, *Euclidean decompositions of noncompact hyperbolic manifolds*, J. Diff. Geom. **27** (1988), 67–80.

[18] B. Farb − D. Margalit, "A primer on the mapping class group," Princeton Mathematical Series **49**, Princeton University Press, 2011.

[19] A. Fathi − F. Laudenbach − V. Poénaru, "Travaux de Thurston sur les surfaces," Astérisque, **66**-**67**, Soc. Math. France, Paris, 1979.

[20] A. T. Fomenko − S. V. Matveev, "Algorithmic and computer methods for three-manifolds", Mathematics and its Applications, **425**, Kluwer Academic Publishers, Dordrecht, 1997.

[21] D. Futer − F. Guéritaud, *From angled triangulations to hyperbolic structures*, Contemporary Mathematics **541** (2011), 159–182.

[22] D. Gabai − R. Meyerhoff − P. Milley, *Minimum volume cusped hyperbolic three-manifolds*, J. Amer. Math. Soc. **22** (2009), 1157–1215.

[23] R. E. Gompf − A. I. Stipsicz, "4-Manifolds and Kirby calculus," Amer. Math. Soc. Providence, RI (1999).

[24] U. Haagerup − H. J. Munkholm, *Simplices of maximal volume in hyperbolic n-space*, Acta Math. **147** (1981), 1–11.

[25] A. Hatcher, "Algebraic Topology," Cambridge University Press, 2002, available from `https://www.math.cornell.edu/ hatcher/AT/ATpage.html`

[26] _____, "Notes on Basic 3-Manifold Topology," available from `http://www.math.cornell.edu/~hatcher/3M/3Mdownloads.html`

[27] _____, *Measured lamination spaces for surfaces from the topological viewpoint*, Topology Appl. **30** (1988), 63–88.

[28] C. D. Hodgson − J. R. Weeks, *Symmetries, isometries and length spectra of closed hyperbolic three-manifolds*, Experiment. Math. **3** (1994), 261-274.

[29] J. Hoste − M. Thistlethwaite − J. Weeks, *The First 1,701,936 Knots*, Math. Intell. **20**, 33–48.

[30] J. H. Hubbard, "Teichmüller Theory and Applications to Geometry, Topology, and Dynamics," Volume I: Teichmüller Theory.

[31] W. H. Jaco − P. B. Shalen, *Seifert fibered spaces in 3-manifolds*, Memoirs of Amer. Math. Soc. **21** (1979).

[32] K. Johannson, "Homotopy equivalences of 3-manifolds with boundaries," Lecture Notes in Mathematics **761**, Springer, Berlin (1979)

[33] H. Kneser, *Geschlossene Flächen in dreidimensionalen Mannigfaltigkeiten*, Jahr. Deutsch. Math. Verein. **38** (1929), 248–260.

[34] A. Kosinski, "Differential Manifolds", Academic Press, 1993.

[35] J. Lee, "Introduction to Riemannian Manifolds", Springer Graduate Texts in Mathematics **176**, 2018.

[36] D. D. Long, *Immersions and embeddings of totally geodesic surfaces*, Bull. London Math. Soc. **19** (1987), 481–484.

[37] F. Luo − S. Schleimer − S. Tillmann, *Geodesic ideal triangulations exist virtually*, Proc. Amer. Math. Soc. **136** (2008), 2625–2630.

[38] B. Martelli − C. Petronio, *Dehn filling of the 'magic' 3-manifold*, Comm. Anal. Geom. **14** (2006), 967–1024.

[39] C. McMullen, "Riemann surfaces, dynamics and geometry," available from `http://www.math.harvard.edu/~ctm/papers/index.html`

[40] _____, *The Gauss-Bonnet theorem for cone manifolds and volumes of moduli spaces*, to appear in Amer. J. Math.

[41] J. Milnor, *A unique factorization for 3-manifolds*, Amer. J. Math. **84** (1962), 1–7.

[42] W. Neumann − D. Zagier, *Volumes of hyperbolic three-manifolds*, Topology **24** (1985), 307–332.

[43] P. Orlik, "Seifert manifolds," Lecture notes in mathematics, Springer **291** (1972).

[44] R. C. Penner, *A construction of pseudo-Anosov homeomorphisms*, Trans. Amer. Math. Soc. **310** (1988), 179–197.

[45] G. Perelman, *The entropy formula for the Ricci flow and its geometric applications*, 2002, `arXiv:math/0211159`

[46] _____, *Ricci flow with surgery on three-manifolds*, 2003, `arXiv:math/0303109`

[47] _____, *Finite extinction time for the solutions to the Ricci flow on certain three-manifolds*, 2003, `arXiv:math/0307245`

[48] C. Petronio − J. Porti, *Negatively oriented ideal triangulations and a proof of Thurston's hyperbolic Dehn filling theorem*, Expos. Math. **18** (2000), 1–35.

[49] J. Ratcliffe, "Foundations of hyperbolic manifolds", Graduate texts in mathematics **149**, Springer-Verlag, 1994.

[50] D. Rolfsen, "Knots and links," Publish or Perish (1976).

[51] T. B. Rushing, "Topological embeddings," Pure and Applied Mathematics **52**, Academic Press Inc, 1973.

[52] P. Scott, *The geometries of 3-manifolds*, Bull. London Math. Soc. **15** (1983), 401–487.

[53] H. Seifert, *Topologie dreidimensionaler gefaserter Räume*, Acta Math. **60** (1933) 147–238.

[54] S. Smale, *Diffeomorphisms of the 2-sphere*, Proc. Amer. Math. Soc. **10** (1959), 621–626.

[55] W. Thurston, *On the geometry and dynamics of diffeomorphisms of surfaces*, Bull. Amer. Math. Soc. **19** (1988), 417–431.

[56] _____ "The geometry and topology of three-manifolds," Princeton University Mathematics Department, 1979.

[57] _____, "Three-dimensional geometry and topology," Volume I, edited by Silvio Levy, Princeton Mathematical Series, **35**, 1997

[58] _____, *Three-dimensional manifolds, Kleinian groups and hyperbolic geometry*. Amer. Math. Society. Bulletin. **6** (1982), 357–381.
[59] E. B. Vinberg, *Hyperbolic reflection groups*, Russian Math. Surveys **40** (1985), 31–75.
[60] F. Waldhausen, *Eine Klasse von 3-dimensionalen Mannigfaltigkeiten*, Invent. Math. **3** (1967), 308-333
[61] K. Yoshida, *The minimal volume orientable hyperbolic 3-manifold with 4 cusps*, Pacific Math. **266** (2013), 457–476.

Index