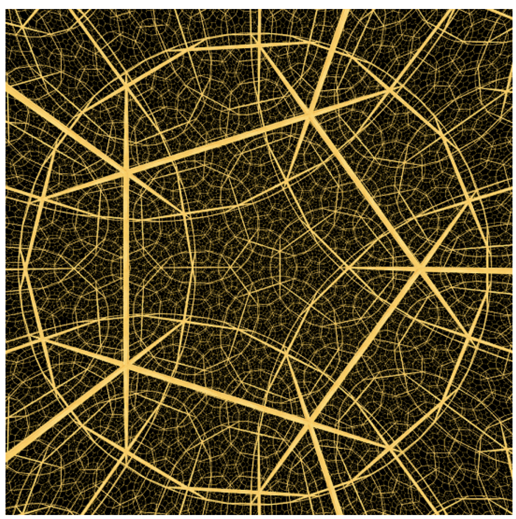
Master's thesis in mathematics

Winter semester 2009-2010

Approaches to Mostow rigidity in hyperbolic space

Adrien Lücker



THE UNIVERSITY OF CHICAGO

Supervised at University of Chicago by :

Prof. Benson Farb
Department of Mathematics

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Responsible professor at EPFL :

Prof. Nicolas Monod
Ergodic and Geometric Group Theory
Department of Mathematics

Abstract

This dissertation presents three proofs of Mostow's Rigidity Theorem. This result states that two compact hyperbolic manifolds of dimension $n \geq 3$ with isomorphic fundamental groups are isometric. Thurston's proof investigates the analytical properties of the boundary map to show that it is conformal. Gromov's proof uses homology theory, Gromov norm and simplices of maximal volume in hyperbolic space. The approach of Besson, Courtois and Gallot gives a characterization of locally symmetric metrics from which Mostow rigidity easily follows. We also present the Dehn-Nielsen-Baer theorem which is an important result the about mapping class group of a surface.

Contents

Introduction 7

Acknowledgements 11

1 Some hyperbolic geometry 13
- 1.1 Basic facts 13
- 1.2 Examples of hyperbolic manifolds 19

2 Mostow's Rigidity Theorem 25
- 2.1 Consequences and outline 26
- 2.2 The boundary map 28
- 2.3 Properties of the boundary map 34
- 2.4 Ergodicity of geodesic flow 39

3 Gromov's proof 47
- 3.1 Gromov norm 47
- 3.2 Simplices and their volume 49
- 3.3 Gromov norm of hyperbolic manifolds 51
- 3.4 Gromov's proof of Mostow rigidity 55

4 Minimal entropy 59
- 4.1 Entropy and outline 60
- 4.2 Patterson-Sullivan measures 64
- 4.3 The barycenter 67
- 4.4 Computation of the Jacobian 70

5 The Dehn-Nielsen-Baer Theorem 75
- 5.1 Change of coordinates principle 77
- 5.2 Quasi-isometries 79

5.3 End of proof 82

Bibliography **85**

Index **89**

Introduction

Mostow's rigidity theorem is seminal in the theory of rigidity. It is one of the most influential results of the last fifty years in geometry and topology. This dissertation attempts not only to explain and prove it, but also to illustrate the wide range of mathematical techniques this theorem is linked to. Thus, rather than simply proving Mostow rigidity in the simplest possible way, we give three proofs of it.

Assume that two Riemannian manifolds have isomorphic fundamental groups. From a geometer's point of view, this is a topological, unprecise information that does not bring much geometric information. However, under the hypotheses of Mostow's rigidity theorem, having the same fundamental group implies being isometric ! Precisely,

Theorem. *Let M, N be compact hyperbolic n-manifolds, $n \geq 3$. If $\pi_1(M)$ is isomorphic to $\pi_1(N)$, then M and N are isometric. Moreover, the isomorphism of fundamental groups is induced by a unique isometry.*

Moreover, this theorem is false in dimension 2 (see the beginning of Chapter 2). The proofs that we are going to present include tools such as the boundary at infinity, quasi-isometries, conformal maps, some analysis results, ergodic theory, homology theory and differential geometry.

Chapter 1 collects basic facts that will be needed subsequently and provides some examples of hyperbolic manifolds. Although it is not necessary to read it, its presence is meant to emphasize that it is a good idea to have some intuition in hyperbolic geometry before jumping into proofs of Mostow rigidity.

The common denominator of the next chapters is a construction known as the boundary map. Whereas the frequent use of universal covers should not be a surprise, the fact that much of the work will be carried out on the boundary at infinity of the universal covers of our manifolds is perhaps more interesting. This construction is given in Section 2.2.

Besides the construction of the boundary map, Chapter 2 discusses Mostow rigidity and presents a version of Mostow's proof modified by Thurston. Using geometric arguments, hard analytical facts and ergodic theory, one proves that the boundary map is conformal and thus gives rise to an isometry between the manifolds. A proof of the ergodicity of the geodesic flow on finite-measure hyperbolic manifolds is given in Section 2.4.

Chapter 3 presents Gromov's proof of Mostow rigidity. It uses a homological invariant of a manifold known as Gromov norm and the volume of simplices in hyperbolic space. The main result of that chapter is that Gromov norm and the volume of hyperbolic manifolds are proportional. Therefore one first proves that the manifolds M and N have the same volume. The end of the proof studies the effect of the boundary map on simplices of maximal volume.

Chapter 4 focuses on the approach by Besson, Courtois and Gallot. They use the entropy of a metric to characterize the locally symmetric metrics among all negatively curled metrics on manifolds related by a homotopy equivalence. Mostow rigidity then follows as an easy corollary. This method mainly uses techniques from differential geometry.

Chapter 5 deals with the Dehn-Nielsen-Baer Theorem. Surfaces are outside the scope of Mostow rigidity because of the hypothesis on dimension. However, hyperbolic geometry and quasi-isometries can be used to prove this result about surfaces which is closely related to a corollary of Mostow rigidity. Let Σ_g be a surface of genus $g \geq 1$. The Dehn-Nielsen-Baer Theorem states that the group of outer automorphisms of $\pi_1(\Sigma_g)$ is isomorphic to the generalized mapping class group of Σ_g.

This dissertation was written so as to be fairly self-contained and accessible to students who have studied mathematics for about four years. Our philosophy was to give enough information and hints so that the reader can fill in the last details by himself. Our hope is to transmit some of the beauty of this subject where many areas of mathematics meet. Concerning prerequisites, instead of writing a list of concepts that should be known before reading this, we suggest the reader to browse through Chapter 1 and to find out by himself how comfortable he feels with the vocabulary.

Acknowledgements

I express my deep gratitude to Prof. Benson Farb for welcoming me to the University of Chicago and guiding me through this Master's thesis. His sheer enthusiasm was a constant source of motivation and my appreciation of the subject owes much to his ability to convey the beauty of mathematics.

I thank Prof. Nicolas Monod for having suggested and made possible this great semester that I spent in Chicago. I am also grateful to the Swiss Federal Institute of Technology of Lausanne for its financial support to academic exchange students.

I thank Katie Mann for her help, the fruitful discussions that we had and her numerous useful remarks about this dissertation.

The picture on the title page shows a tiling of the hyperbolic 3-space by dodecahedra, represented in the Klein model. The image file was taken from Wikimedia Commons and was designed using the Curved Spaces software.

Chapter 1

Some hyperbolic geometry

This chapter gathers facts about hyperbolic geometry that will be used often repeatedly in this dissertation. Readers familiar with the topic should skip this and return only if necessary.

Since most results presented in this work involve hyperbolic manifolds, we devote Section 1.2 to constructions of some of the classical examples.

Throughout this dissertation, we will only consider complete connected orientable manifolds. This will be not be repeated throughout. However, these hypotheses will be written out in the main results' statements. Also, hyperbolic will always mean real hyperbolic and manifolds are without boundary (unless otherwise indicated).

1.1 Basic facts

In this section, we quickly introduce some models of hyperbolic geometry, the boundary at infinity (which will be ubiquitous in this thesis), the classification of isometries in hyperbolic space, conformal maps and Busemann functions. Only a few proofs are given, so the reader who is new to these concepts should consult other references. This material has been gathered here so as to make this thesis more self-contained, but not to provide an actual introduction to hyperbolic geometry.

Hyperboloid model

We first define hyperbolic space using the hyperboloid model as follows:

$$\mathbb{H}^n = \{(x_1, \dots, x_{n+1}) \in \mathbb{R}^{n+1} : x_1^2 + \dots + x_n^2 - x_{n+1}^2 = -1,\ x_{n+1} > 0\}.$$

For $p \in \mathbb{H}^n$, the restriction of the Euclidean scalar product on $\mathbb{R}^{n+1}$ to $T_p\mathbb{H}^n$ is positive definite. This endows $\mathbb{H}^n$ with a Riemannian metric. In this model, the isometries of $\mathbb{H}^n$ are given by

$$\mathrm{Isom}(\mathbb{H}^n) = \{M = (m_{ij}) \in \mathrm{GL}(n+1, \mathbb{R}) : M^t I_{n,1} M = M,\ m_{n+1,n+1} > 0\},$$

where

$$I_{n,1} = \begin{pmatrix} 1 & & & \\ & \ddots & & \\ & & 1 & \\ & & & -1 \end{pmatrix}.$$

The group $\mathrm{Isom}(\mathbb{H}^n)$ acts $(n+1)$-transitively on $\mathbb{H}^n$. This means that for points $p,\ q \in \mathbb{H}^n$ and orthonormal frames at these points, there exists an isometry of $\mathbb{H}^n$ mapping one orthonormal frame to another. Besides, it is an exercise in Riemannian geometry to show that the curve $t \mapsto (0, \dots, 0, \sinh t, \cosh t)$ is a geodesic. Using these facts, it is easy to see that any geodesic is given by the intersection of an (Euclidean) hyperplane through the origin and $\mathbb{H}^n$. Hyperbolic hyperplanes are totally geodesic complete submanifolds of $\mathbb{H}^n$. This means that geodesics in the hyperplane are geodesics in $\mathbb{H}^n$ and that they can be infinitely extended within the hyperplane.

Ball model

Another model of hyperbolic space can be obtained by stereographic projection of the hyperboloid onto the unit sphere in $\mathbb{R}^n$, based at the point $(0, \dots, 0, -1)$. This model is especially useful for visualizing $\mathbb{H}^3$. Geodesics are Euclidean circles cutting orthogonally the boundary S^{n-1} of the unit n-ball. Likewise, hyperbolic hyperplanes come from Euclidean spheres cutting S^{n-1} orthogonally.

Upper half-space model

A third model is obtained by inverting the unit ball model through a sphere of radius $\sqrt{2}$ centered at $(0, \ldots, 0, -1)$. This maps the unit ball onto the upper half-space

$$\{(x_1, \ldots, x_n) \in \mathbb{R}^n : x_n > 0\}.$$

Geodesics are either Euclidean semi-circles orthogonally cutting the horizontal plane or vertical lines. Similarly, hyperplanes are Euclidean hemispheres meeting the horizontal plane orthogonally or vertical hyperplanes. This model is also useful for visualization purposes and makes some calculations easier.

The boundary at infinity

It would not be too big an exaggeration to state that we will encounter maps on the boundary at infinity of $\mathbb{H}^n$ more often than maps on the hyperbolic space itself. The three proofs of Mostow rigidity expounded in this thesis will involve a map $\tilde{f} : \mathbb{H}^n \to \mathbb{H}^n$ which is the lift of a homotopy equivalence between compact hyperbolic manifolds. Our goal is to straighten this lift to an isometry. This will always be done by studying (in various ways) the effect of $\tilde{f}$ on the boundary at infinity of $\mathbb{H}^n$.

Let X be a nonpositively curled manifold. A ***geodesic ray*** is a geodesic $\varphi : [0, \infty) \to X$. Two geodesic rays φ_1 and φ_2 are ***asymptotic*** if the distance between $\varphi_1(t)$ and $\varphi_2(t)$ stays bounded for all $t \geq 0$. This defines an equivalence relation on the set of geodesic rays in X, denoted by $\sim$. The ***boundary at infinity*** of X is defined to be

$$\partial X = \{\text{geodesic rays in } X\}/\sim .$$

The equivalence class of a geodesic ray φ is denoted by $\varphi(\infty)$. We also write

$$\bar{X} = X \cup \partial X.$$

We endow the space $\bar{X}$ with the ***cone topology***. Fix a basepoint $x_0 \in X$. A basis of neighborhoods is given by open sets in X and sets of the form

$$\{y \in \bar{X} : \angle(y, \xi) < \varepsilon,\, d(x_0, y) > r\} \quad \text{for } \varepsilon > 0,\, r > 0,\, \xi \in \partial X.$$

This topology does not depend on the choice of basepoints. It is easy to see that in the ball model, $\overline{\mathbb{H}^n}$ is homeomorphic to the closed unit ball in $\mathbb{R}^n$.

Also, since isometries send open basis sets to themselves, an isometry of X naturally extends to a homeomorphism of $\bar{X}$.

For the universal cover $\tilde{X}$ of a complete nonpositively curled manifold X, the boundary at infinity can be identified with S^{n-1} by choosing any $x \in \tilde{X}$ and "looking around". More precisely, for $x \in \tilde{X}$, the map

$$\begin{array}{rcl} T^1_x\tilde{X} & \longrightarrow & \partial\tilde{X} \\ v & \longmapsto & \gamma_v(\infty) \end{array}$$

is a homeomorphism (as usual γ_v denotes the geodesic with initial tangent vector v). Very often, S^{n-1} will be implicitly identified with $\partial\tilde{X}$.

Transitivity properties of the group of isometries of $\mathbb{H}^n$ often allow to make simplifying assumptions. This will sometimes be used without reference.

Proposition 1.1.

1. $\mathrm{Isom}(\mathbb{H}^n)$ *acts transitively on the set of k-hyperplanes,* $1 \leq k \leq n-1$.
2. $\mathrm{Isom}(\mathbb{H}^n)$ *acts n-transitively on the boundary at infinity of* $\mathbb{H}^n$.
3. *In the upper half-space model, the set of isometries fixing the point at infinity equals the group of Euclidean similarities.*

Classification of isometries

The vocabulary we now introduce will be used repeatedly. Isometries of hyperbolic space split into three categories. It is useful to know two equivalent definitions. A nontrivial isometry $\gamma \in \mathrm{Isom}(\mathbb{H}^n)$ is

- ***elliptic*** if it fixes a unique point in $\mathbb{H}^n$;
- ***parabolic*** if it fixes a unique point at infinity;
- ***hyperbolic*** if it fixes two points at infinity.

For the alternate definition, define the displacement function of an isometry γ of $\mathbb{H}^n$ to be

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

Then γ is

- ***elliptic*** if $D_\gamma = 0$ and the infimum is attained;
- ***parabolic*** if $D_\gamma = 0$ and the infimum is not attained;
- ***hyperbolic*** if $D_\gamma > 0$ and the infimum is attained.

For the former and the latter definition, it is an exercise to show that all the cases are exhausted and that they are equivalent. See [23, Prop. 2.5.17] for a complete proof.

It is useful to understand why hyperbolic isometries are going occur most often in this dissertation. Indeed, complete hyperbolic manifolds of finite area arise from quotients of $\mathbb{H}^n$ by a lattice Γ in $\mathrm{Isom}(\mathbb{H}^n)$. To obtain a manifold, the action of Γ on $\mathbb{H}^n$ should have no fixed points, so Γ should contain no elliptic elements.

Whereas fixed points rule out elliptic isometries, compactness exludes parabolic isometries. Let $\gamma \in \Gamma$. Since the function $x \mapsto d(x, \gamma.x)$ is continuous in hyperbolic space, compactness of the manifold implies that the infimum must be attained and so γ cannot be parabolic. Thus γ must be hyperbolic. As we will primarily focus on compact hyperbolic manifods, we will mainly encounter hyperbolic isometries. This will be illustrated in the examples given in Section 1.2.

In fact, the condition that Γ should not contain elliptic isometries can be rephrased in a convenient way, namely that Γ has to be torsion-free. Assume that an element γ has torsion. This means that the orbits of γ are finite. In nonpositive curvature, there is a meaningful concept of center of mass of a bounded set (see [6, Prop. 2.7]. The center of mass of that finite orbit must be fixed by γ, but γ should have no fixed points. This discussion proves most of the following result. It will be used in Chapter 2 to get an algebraic version of Mostow's rigidity theorem.

Proposition 1.2. *There is a natural bijection*

$$\{\textit{isometry classes of complete hyperbolic manifolds of finite volume}\}$$
$$\leftrightarrow$$
$$\{\textit{conjugacy classes of torsion-free lattices in } \mathrm{Isom}(\mathbb{H}^n)\}.$$

Moreover, compact manifolds correspond to cocompact lattices.

Proof. There just remains to see that such manifolds are isometric if and only if the underlying lattices are conjugate. This is an easy exercise. □

Conformal maps and inversions

The interplay between isometries of $\mathbb{H}^n$ and maps on the boundary S^{n-1} will be crucial in Chapter 2. We mentioned that isometries of $\mathbb{H}^n$ induce homeomorphisms of $\overline{\mathbb{H}^n}$. Knowing what kind of boundary maps give rise to isometries of $\mathbb{H}^n$ will be essential. Here is the kind of maps that will arise.

Definition 1.3. Let X, Y be Riemannian manifolds. A diffeomorphism $f : X \to Y$ is ***conformal*** if there exists a smooth function $\alpha : X \to \mathbb{R}_+$ such that for all $x \in X$,

$$\langle Df(u), Df(v)\rangle_{f(p)} = \alpha(p)\langle u, v\rangle_p \quad \text{for all } u, v \in T_pX.$$

Notice that conformality of f is equivalent to the fact that f preserves angles.

We are interested in conformal maps because they arise naturally from isometries of $\mathbb{H}^n$. One can prove that any isometry of $\mathbb{H}^n$ is a composition of reflections through $(n-1)$-hyperplanes (see [6, Prop. 2.17]). Using the ball model, it is easy to see that symmetries of $\mathbb{H}^n$ are in one-to-one correspondence with inversions of S^{n-1}.

Theorem 1.4. *Denote by* $\mathrm{Conf}(S^{n-1})$ *the group of conformal maps of* S^{n-1}*. For* $n \geq 2$*, the natural map*

$$\mathrm{Isom}(\mathbb{H}^n) \longrightarrow \mathrm{Conf}(S^{n-1})$$

is an isomorphism.

Proof. Well-definedness and injectivity were explained above. To prove surjectivity, one needs that for $n \geq 2$, the conformal group of S^{n-1} is generated by inversions. This is a nontrivial fact which can be found in [2]. □

Busemann functions

This paragraph will only be needed in Chapter 4, so it is recommended to skip it until it is needed. Busemann functions formalize the notion of distance to infinity. In the following, we assume X to be a nonpositively curved manifold and φ a geodesic ray in X. Let also $\xi = \varphi(\infty)$.

Definition 1.5. For $x_0 = \varphi(0)$, the ***Busemann function at*** x_0 ***centered at*** ξ is defined to be

$$B_{x_0}(x, \xi) = \lim_{t \to \infty} (d(x, \varphi(t)) - t).$$

Busemann functions behave nicely with respect to change of basepoints. More precisely, for x_1, $x_2 \in X$ and $\xi \in \partial X$ as above, one has

$$B_{x_0}(\cdot, \xi) - B_{x_1}(\cdot, \xi) = C,$$

where the constant C depends on x_0 and x_1.

Horospheres are level sets of Busemann functions. More precisely, the ***horosphere at*** ξ ***through*** x_1 is the set

$$HS(x_1, \xi) = \{x \in X : B_{x_1}(x, \xi) = 0\}.$$

To visualize a Busemann function B_{ξ, x_0}, it is easiest to work in the upper half space model, putting ξ at ∞. Then horospheres are horizontal planes. The Busemann function $B_{x_0}(x_1, \xi)$ is the signed distance between horospheres through x_0 and x_1 respectively. In this case, it is the vertical hyperbolic distance between the two horizontal planes. If ξ lies on the horizontal boundary hyperplane, then any horosphere at ξ is a Euclidean sphere tangent to the horizontal plane at ξ. In the ball model, horospheres are also Euclidean spheres tangent to the boundary of the unit ball.

1.2 Examples of hyperbolic manifolds

In this section, we shortly introduce two kinds of constructions of hyperbolic manifolds. First, we show how to put a hyperbolic metric of finite area on all genus $g \geq 1$ punctured surfaces, except the torus that does not admit a hyperbolic structure. This is done by directly constructing fundamental domains in $\mathbb{H}^2$. Next, we provide a construction of hyperbolic arithmetic manifolds that provides less geometric intuition but is very general, although we introduce one example only.

Hyperbolic surfaces

The canonical example is provided by Σ_g, the closed oriented genus g surface, when $g \geq 2$. Recall that Σ_g is obtained by gluing a $4g$-gon with edges labelled $a_1, b_1, a_1^{-1}, b_1^{-1} \dots, a_g, b_g, a_g^{-1}, b_g^{-1}$ via the gluing pattern $a_1 b_1 a_1^{-1} b_1^{-1} \dots a_g b_g a_g^{-1} b_g^{-1}$. Since the sum of the angles of a Euclidean $4g$-gon is $(4g-2)\pi$, the resulting CW complex cannot be embedded in $\mathbb{R}^3$ when $g \geq 2$. Contrary to this, flat tori can be embedded in $\mathbb{R}^3$ since they arise from a quadrilateral whose angles sum up to 2π.

Observe that one can construct hyperbolic $4g$-gons whose angles sum up to 2π. To see this, use the ball model (where Euclidean angles equal hyperbolic angles). A tiny $4g$-gon is "almost Euclidean", whereas a very large $4g$-gon has angles close to 0. By continuity, some scaling of a $4g$-gon has angles whose sum is 2π. To obtain a hyperbolic manifold of the form $\mathbb{H}^2/\Gamma$, the gluing has to be done via isometries. The next elementary result ensures that such isometries exist.

Proposition 1.6. *Given two distinct geodesics γ_1,γ_2 in $\mathbb{H}^n$, there exists a unique geodesic φ in $\mathbb{H}^n$ that crosses orthogonally γ_1 and γ_2.*

Sketch of proof. Work in the ball model. The two geodesics are contained in an embedded copy of $\mathbb{H}^2$. Then use a continuity argument involving the set of geodesics cutting orthogonally γ_1. □

Thus, to endow the surface Σ_g, $g \geq 2$ with a hyperbolic metric, construct such a $4g$-gon in $\mathbb{H}^2$ and glue its sides via the pattern described above. Figure 1.1 shows a regular octagon with angles equal to $\pi/4$. The dashed lines are the axes of hyperbolic isometries mapping one side to another.

Some manifolds do not admit a hyperbolic metric. Recall that if S a closed surface, the Gauss-Bonnet theorem implies that

$$\int_S K = 2\pi\chi(S),$$

where K is the Riemann curvature of S and $\chi(S)$ is the Euler characteristic of S. In particular, the sphere and the torus (of Euler characteristic 1 and 0 resp.) do not admit a hyperbolic metric.

Concerning tori, the situation is different when they are punctured. Denote by $\Sigma_{g,p}$ the genus g surface with p punctures. We are going to describe how to put a complete hyperbolic metric of finite area and infinite diameter on $\Sigma_{g,p}$ for all $g \geq 1$, $p \geq 1$.

Let us start with $\Sigma_{1,1}$. A quadrilateral Q in $\mathbb{H}^2$ is said to be ideal if all its vertices lie on the boundary at infinity of $\mathbb{H}^2$. By Proposition 1.6, there exists geodesics γ, ρ cutting orthogonally pairs of opposites sides of Q. These

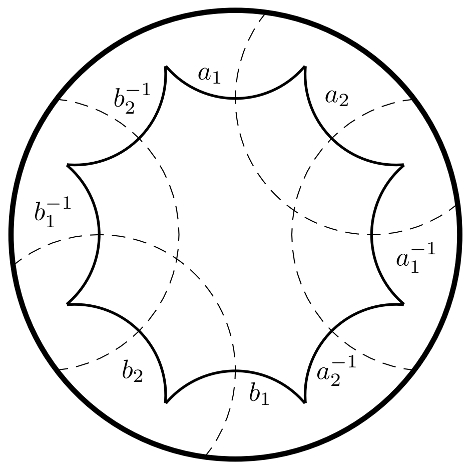


Figure 1.1: Construction of a hyperbolic structure on Σ_2

geodesics give rise to hyperbolic isometries g and h that both map a side of Q to its opposite side. Let $\Gamma = \langle g, h \rangle$. It is easy to see that Γ is the free group on two elements, which is incidentally the fundamental group of $\Sigma_{1,1}$ (observe that $\Sigma_{1,1}$ deformation retracts to a wedge of two circles). Gluing opposite sides of Q yields a 1-punctured torus. Although the generators of Γ are both hyperbolic, Γ contains parabolic elements, for instance $ghg^{-1}h^{-1}$. This is consistent with the discussion on page 17 and the fact that $\Sigma_{1,1}$ is noncompact.

This construction easily generalizes. To construct $\Sigma_{2,p}$, start with an octagon whose vertices lie on the boundary of $\mathbb{H}^2$ and add $p-1$ "bumps" to the sides a_1 and a_1^{-1} as shown on Figure 1.2. Join again these bumps with geodesic lines cutting them orthogonally and glue the polygon accordingly. It is a good exercise for visualization to observe that this yields $\Sigma_{2,p}$.

Arithmetic constructions

Arithmetic lattices are an essential (and huge) class of lattices. By the way, their importance reaches far beyond hyperbolic geometry. Intuitively, arithmetic lattices are obtained by "taking integer points" in a semisimple Lie group. The canonical example is $\mathrm{SL}(n, \mathbb{Z})$, which is a lattice in $\mathrm{SL}(n, \mathbb{R})$.

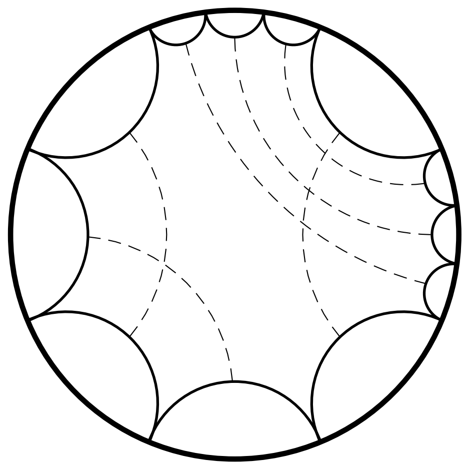


Figure 1.2: Construction of a hyperbolic structure on $\Sigma_{2,3}$

However, in order to have a reasonable concept arithmetic lattice should be stable under some basic operations, such as isomorphisms, modding out by compact subgroups and passing to finite index subgroups. This is why the formal definition of arithmetic lattices is somewhat lengthy – we just mention to satisfy the curiosity of the interested reader. We will not explain all the vocabulary used in the definition.

Definition 1.7. Let Γ be a lattice in a semisimple Lie group G. Γ is an ***arithmetic lattice*** in G if there exists

- a closed connected semisimple subgroup G' of some $\mathrm{SL}(l, \mathbb{R})$ that is defined over $\mathbb{Q}$,
- compact normal subgroups $K < G$ and $K' < G'$,
- an isomorphism $\varphi : G/K \to G'/K'$,

such that $\varphi(\overline{\Gamma})$ is commensurable with $\overline{G'_{\mathbb{Z}}}$. Here $\overline{\Gamma}$ and $\overline{G'_{\mathbb{Z}}}$ are the images of Γ and $G'_{\mathbb{Z}}$ in G/K and G'/K' respectively.

For more information on arithmetic lattices, the reader should consult [15].

The following tool is useful to show that spaces of lattices (and thus manifolds) are compact. Given a lattice Λ in $\mathbb{R}^l$, we define $\min(\Lambda) = \min\{\|v\| :$

$v \in \Lambda\}$, where $\|\cdot\|$ is the Euclidean norm.

Theorem 1.8. Mahler Compactness Criterion. *Let L be a space of covolume 1 lattices. Then L is precompact if and only if there exists $\varepsilon > 0$ such that $\min(\Lambda) \geq \varepsilon$ for all $\Lambda \in L$.*

Example 1.9. Cocompact arithmetic lattice. Define quadratic forms

$$\begin{matrix} h_-(x) = x_1^2 + \ldots + x_n^2 - \sqrt{2}x_{n+1}^2 \\ h_+(x) = x_1^2 + \ldots + x_n^2 + \sqrt{2}x_{n+1}^2 \end{matrix} \qquad \text{for } x = (x_1, \ldots, x_{n+1}) \in \mathbb{R}^{n+1}.$$

Recall that given a Lie group H, we denote the connected component of the identity by H°. Let $G = \mathrm{SO}(h_-)^\circ$ and let $\Gamma = G \cap \mathrm{SL}(n+1, \mathbb{Z}[\sqrt{2}])$ We are going to show that Γ is a cocompact lattice in G. At first sight, it is not even obvious that Γ is discrete in G (indeed, $\mathbb{Z}[\sqrt{2}]$ is not discrete in $\mathbb{R}$).

For simplicity, write $R = \mathbb{Z}[\sqrt{2}]$. To show discreteness of Γ, let σ be the (only) nontrivial Galois automorphism of R given by $\sigma(\sqrt{2}) = -\sqrt{2}$. Let $\Delta = (\mathrm{id}, \sigma)$. Since $\Delta(1) = (1,1)$ and $\Delta(\sqrt{2}) = (\sqrt{2}, -\sqrt{2})$ are linearly independent over $\mathbb{R}$, it follows that $\Delta(\Gamma)$ is discrete in $\mathrm{GL}(n+1,\mathbb{R}) \times \mathrm{GL}(n+1, \mathbb{R})$. An elementary calculation shows that $\Delta(\Gamma) \subset G \times \mathrm{SO}(n+1)$. Since the second factor is compact, the image of Γ in the first factor must be discrete. Hence Γ is discrete in G.

We use the fact that $G/\operatorname{Stab}_G(x_0)$ is homeomorphic to $G.x_0$. In our situation, the lattice $\Delta(R^{n+1})$ of $\mathbb{R}^{2(n+1)}$ plays the role of x_0. Thus

$$\Delta(\Gamma) = \operatorname{Stab}_{G\times \mathrm{SO}(n+1)}(\Delta(R^{n+1})) \simeq (G \times \mathrm{SO}(n+1)).(\Delta(R^{n+1})).$$

We claim that the space of lattices $(G \times \mathrm{SO}(n+1)).\Delta(R^{n+1})$ has a positive minimum. Another elementary calculation shows that the value of the function $h(x) := h_-(x)h_+(\sigma(x))$ is an integer for all $x \in R^{n+1}$. The neighborhood of 0 in $\mathbb{R}^{n+1} \times \mathbb{R}^{n+1}$ defined by $\{(u,v) \in \mathbb{R}^{n+1} \times \mathbb{R}^{n+1} : |h_-(u)h_+(v)| < 1\}$ contains no other elements of $\Delta(\Gamma)$. Indeed, if $h_-(x)h_+(\sigma(x)) = 0$ for some $x \in R^{n+1}$, then $h_-(x)$ must be zero. Thus $h_+(\sigma(x)) = 0$, and so $x = 0$. This proves the claim. It now follows immediately that $G.R^{n+1}$ has a positive minimum. By Mahler's compactness criterion 1.8, $G.R^{n+1}$ must be precompact.

It remains to show that G/Γ is closed. Let $(\Lambda_r)_{r=1}^{\infty}$ be a sequence of lattices in $(G \times \mathrm{SO}(n+1)).\Delta(R^{n+1})$ converging to some lattice Λ in $\mathbb{R}^{n+1} \times \mathbb{R}^{n+1}$. The values of h on vectors of Λ_i are contained in a discrete set. For an index

i_0 sufficiently large, the function h assume the same values on some ordered bases of Λ_{i_0} and Λ. This means that there is an element of $G \times \mathrm{SO}(n+1)$ mapping Λ_{i_0} to Λ. In particular, the first factor of Λ is in $G.R^{n+1}$. Thus $G/\Gamma \simeq G.R^{n+1}$ is both precompact and closed. This proves that G/Γ is compact.

A last detail has to be taken care of. Nothing tells us that Γ is torsion-free. Nevertheless, the following result asserts that a subgroup of Γ of finite index has that desired property.

Theorem 1.10. Selberg's lemma. *Let G be a linear group and $\Gamma < G$ a subgroup of finite index. Then Γ has a torsion-free subgroup of finite index.*

Thus, Selberg's lemma provides us with a torsion free subgroup $\Lambda < \Gamma$ of finite index and thus we obtain a hyperbolic manifold G/Λ.

Chapter 2

Mostow's Rigidity Theorem

As mentioned in the introduction, this chapter presents Thurston's proof of Mostow's rigidity theorem. This theorem was originally proved in 1967 by the American mathematician George Mostow in [16]. Some of its ideas are present in Thurston's proof but the two proofs remain essentially different.

Theorem 2.1. Mostow rigidity. *Let M, N be compact hyperbolic n-manifolds with $n \geq 3$. Assume that M and N have isomorphic fundamental groups. Then the isomorphism of fundamental groups is induced by a unique isometry.*

Besides this geometric version, there is an equivalent algebraic version of Mostow rigidity. The proof of the equivalence is a simple application of Proposition 1.2, which translates information about manifolds into the language of lattices.

Theorem 2.2. Mostow rigidity, algebraic version. *Let Γ_1, Γ_2 be co-compact lattices in* $\mathrm{Isom}(\mathbb{H}^n)$, *with $n \geq 3$. If they are isomorphic, then they are conjugate in* $\mathrm{Isom}(\mathbb{H}^n)$.

Let us introduce a definition to attempt to show the significance of Mostow's rigidity theorem.

Definition 2.3. A lattice Γ in a Lie group G is said to be ***strongly rigid*** or ***Mostow rigid*** if, given a lattice Γ' in another Lie group G' and an isomorphism $\varphi : \Gamma \to \Gamma'$, φ extends to a unique isomorphism $G \to G'$.

It follows the algebraic version of Mostow's rigidity theorem that we have strong rigidity when $G = G' = \mathrm{Isom}(\mathbb{H}^n)$, $n \geq 3$ and Γ and Γ' are fundamental groups of compact hyperbolic manifolds. The problem of rigidity of lattices in Lie groups is very general. Theorems 2.1 and 2.2 were the first results of this kind and many rigidity theorems were discovered after Mostow's breakthrough.

Throughout this dissertation, we will deal with the geometric version of Mostow's rigidity. Also, we will focus on compact manifolds. Nevertheless, Prasad showed that the same statement holds for finite-volume complete hyperbolic manifolds (see [19]).

Before going into corollaries and the proof of Mostow rigidity, let us discuss one of the hypotheses. The assumption that $n \geq 3$ is essential. In the three proofs that will be given, this hypothesis will be used in very different ways. To find an elementary counter-example to Mostow rigidity in dimension 2, find two octagons in $\mathbb{H}^2$ whose sum of angles is 2π and that are not isometric. They form fundamental domains for nonisometric surfaces of genus 2.

In fact, Teichmüller theory tells us that the space of all marked hyperbolic structures on Σ_g is homeomorphic to $\mathbb{R}^{6g-6}$ (see [7, Chapter 9]). Therefore, such manifolds can be deformed and are not rigid. The whole point of Mostow rigidity is that this kind of deformations cannot happen in higher dimensions.

2.1 Consequences and outline

We now mention some consequences of Mostow rigidity. Assume that two compact hyperbolic manifolds M and N are homeomorphic. A priori, there is no reason for M and N to have the same geometric invariants (such as volume, diameter, injectivity radius). However, Mostow rigidity implies that this actually happens.

Corollary 2.4. *For compact hyperbolic manifolds, geometric invariants are topological invariants.*

Proof. The manifolds M and N are homeomorphic so they have isomorphic fundamental groups. Therefore they are isometric and have the same geometric invariants. □

Given a group G, the group of outer automorphisms of G is defined to be $\mathrm{Out}(G) = \mathrm{Aut}(G)/\mathrm{Inn}(G)$, where elements of $\mathrm{Inn}(G)$ are inner automorphisms of G. For a fixed basepoint $x_0 \in M$, any isometry φ of M induces an automorphism

$$\varphi_* : \pi_1(M, x_0) \longrightarrow \pi_1(M, \varphi(x_0)).$$

In order to get an automorphism of $\pi_1(M)$, we have to choose a path from x_0 to $\varphi(x_0)$. Thus we do not obtain a canonical map $\mathrm{Isom}(M) \to \mathrm{Aut}(\pi_1(M))$. However, observing that choosing different paths results in a conjugation by an element of $\pi_1(M, x_0)$, modding out by $\mathrm{Inn}(\pi_1(M))$ yields a canonical map from $\mathrm{Isom}(M)$ to $\mathrm{Out}(\pi_1(M))$.

Corollary 2.5. *Let M be a compact connected orientable hyperbolic n-manifold with $n \geq 3$. Then the natural map*

$$\begin{array}{rcl} \mathrm{Isom}(M) & \longrightarrow & \mathrm{Out}(\pi_1(M)) \\ F & \longmapsto & [F_*] \end{array}$$

is an isomorphism. Therefore $\mathrm{Out}(\pi_1(M))$ is finite.

Proof. Only the proof of surjectivity uses Mostow's rigidity. Simply apply the theorem with $N = M$. To prove injectivity, let φ be a map inducing the identity outer automorphism of $\pi_1(M)$. Lift φ to an isometry $\tilde{\varphi}$ of $\mathbb{H}^n$. Then $\tilde{\varphi}$ is homotopic to the identity map. Homotopies move points by a bounded distance, so $\tilde{\varphi}$ equals the identity of the boundary at infinity of $\mathbb{H}^n$. Theorem 1.4 implies that $\tilde{\varphi}$ is the identity isometry. Thus $\varphi = \mathrm{id}$.

There remains to prove that $\mathrm{Isom}(M)$ is finite. An isometry of M is uniquely determined by its action on an n-frame F_p at some point $p \in M$ via the exponential map. Assume that there are infinitely many isometries of M. By compactness of M, there exist distinct isometries φ, φ' of M sending F_p to frames arbitrarily close to each other. This implies that φ is homotopic to φ'. By the same argument as in the last paragraph, we infer that $\varphi = \varphi'$, which is a contradiction. Therefore, $\mathrm{Isom}(M)$ is finite and so is $\mathrm{Out}(\pi_1(M))$. □

Before plunging into the proof of Mostow rigidity, we first give an outline of Thurston's proof. Assume that M and N are compact hyperbolic n-manifolds, $n \geq 3$, with isomorphic fundamental groups. The first three steps are described in Section 2.2 and give a general construction for a boundary map. This will be used in all three proofs of Mostow rigidity, whereas subsequent steps are specific to Thurston's proof.

1. Since the universal covers $\tilde{M}$ and $\tilde{N}$ are aspherical, there is a homotopy equivalence $f : M \to N$.

2. This map lifts to a $\pi_1(M)$-equivariant quasi-isometry $\tilde{f} : \tilde{M} \to \tilde{N}$.

3. Geometric lemmas show that $\tilde{f}$ induces a homeomorphism $\partial\tilde{f} : \partial\tilde{M} \to \partial\tilde{N}$ that is $\pi_1(M)$-equivariant.

4. The boundary map is quasi-conformal. As a result of a fact of hard analysis, $\partial\tilde{f}$ is differentiable almost everywhere.

5. It follows from the ergodicity of the geodesic flow on compact hyperbolic manifolds, that the action of $\pi_1(M)$ on $S^{n-1} \simeq \partial\tilde{M}$ is ergodic.

6. This together with a geometric argument forces the differential of $\partial\tilde{f}$ to be conformal

7. Another fact from analysis implies that $\partial\tilde{f}$ is a conformal map.

8. Thus $\partial\tilde{f}$ induces an $\pi_1(M)$-equivariant isometry $\tilde{F} : \tilde{M} \to \tilde{N}$ that descends to an isometry $F : M \to N$.

To make the end of the proof look shorter and clearer, steps 5 and 6 will be postponed after step 8.

2.2 The boundary map

The compact hyperbolic manifold M is our main object of interest. However, most of the proof will be carried out on the universal cover $\tilde{M}$ of M. Since $\tilde{M}$ is hyperbolic space, we can use its boundary at infinity. The gist of the proof is to find a $\pi_1(M)$-equivariant map $\tilde{f}$ between the universal covers. Denote by

$$\rho : \pi_1(M) \to \pi_1(N)$$

the isomorphism of fundamental groups. Equivariance means that $f(\gamma.x) = \rho(\gamma).f(x)$, for $x \in \tilde{M}$ and $\gamma \in \pi_1(M) = \pi_1(N)$. This map will induce a homeomorphisms of the boundaries. The study of this boundary map is the core of Thurston's proof. The next result we quote tells us how the universal cover of a hyperbolic manifold looks like.

Theorem 2.6. Hadamard-Cartan. *Let X be a complete manifold with nonpositive sectional curvature. Then the exponential map $\exp : T_pX \to X$ is a covering map for all $p \in X$. In particular, the universal cover $\tilde{X}$ is diffeomorphic to $\mathbb{R}^n$ and X is a $K(\pi, 1)$ space*

Thus, the universal covers $\tilde{M}$ and $\tilde{N}$ are isomorphic to $\mathbb{H}^n$ which is contractible and so the manifolds M and N are $K(\pi, 1)$ spaces, where $\pi = \pi_1(M) \cong \pi_1(N)$. We recall for a group G, a topological space Y is $K(G, 1)$ if $\pi_1(Y) = G$ and $\pi_k(Y) = 0$ for all $k \geq 2$.

Since $K(\pi, 1)$ spaces are uniquely determined up to homotopy equivalence (see [11, Theorem 1B.8]), the manifolds M, N are homotopically equivalent. Let $f : M \to N$ and $g : N \to M$ be continuous maps such that $f \circ g \sim \mathrm{Id}_N$ and $g \circ f \sim \mathrm{Id}_M$. Lift these maps to the universal cover such that the diagram in Figure 2.1 commutes up to homotopy.

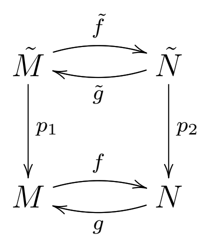


Figure 2.1: Commutativity of f and its lift

We also require another key property of the lifts. The function $\tilde{f}$ can be constructed so that it is $\pi_1(M)$-equivariant. Moreover, by a classical result from differential topology, f and g can be assumed to be C^1.

Since f is C^1, the map $z \mapsto \frac{d(f(x), f(z))}{d(x,z)}$ is continuous. Compactness of M implies that this map is bounded and this means that f is K-Lipschitz for some $K > 0$. As a consequence of the "rubberband principle", $\tilde{f}$ is also Lipschitz. This principle means that being Lipschitz in small scale implies being Lipschitz in large scale. Precisely, pick x and $y \in \tilde{M}$ and subdivide the path between x and y in small neighborhoods such that the covering map p_1 is a local isometry. The same argument implies that g is K-Lipschitz.

We will shortly see that $\tilde{f}$ and $\tilde{g}$ are quasi-isometries, which is defined as follows.

Definition 2.7. Let X, Y be metric spaces and $f : X \to Y$. The map f is said to be a (K, ε)***-quasi-isometry*** if

1. for all x, $x' \in X$,

$$\frac{1}{K} \cdot d(x, x') - \varepsilon \leq d(f(x), f(x')) \leq K \cdot d(x, x') + \varepsilon.$$

2. the map f is coarse onto, that is, there exists $C > 0$ such that every $y \in Y$ lies in a C-neighborhood of some $f(x)$.

Since $\tilde{f}$ and $\tilde{g}$ are K-Lipschitz, they satisfy the upper bound of the last definition. Strictly speaking, no $\varepsilon > 0$ is needed in the upper-bound. This will only matter in the proof of Lemma 2.8. Apart from this, $\tilde{f}$ and $\tilde{g}$ will be treated as usual quasi-isometries.

To prove the lower bound, observe that $g \circ f \sim \mathrm{Id}_M$, so that $\tilde{g} \circ \tilde{f} \sim \mathrm{Id}_{\tilde{M}}$. Since homotopies move points by a bounded distance, there exists $c > 0$ such that $d(\tilde{g} \circ \tilde{f}(x), \tilde{g} \circ \tilde{f}(y)) < c$ for all x, $y \in \tilde{M}$. Using that $\tilde{g}$ is K-Lipschitz, we get

$$\begin{aligned} d(\tilde{f}(x), \tilde{f}(y)) &\geq \frac{1}{K} d(\tilde{g} \circ \tilde{f}(x), \tilde{g} \circ \tilde{f}(y)) \\ &\geq \frac{1}{K}(d(x, y) - 2c). \end{aligned}$$

It remains to see that $\tilde{f}$ is coarse onto. The manifolds M and N are bounded metric spaces, so that f is obviously coarse onto. Then equivariance implies that this must also be true for the lift $\tilde{f}$. Up to a modification of the constants, the same holds for $\tilde{g}$. Thus $\tilde{f}$ and $\tilde{g}$ are (K, ε)-quasi-isometries, with $\varepsilon = 2c/K$.

The next step is to construct a map $\partial \tilde{f} : \partial \tilde{M} \to \partial \tilde{N}$. Recall that $\partial \tilde{M}$ and $\partial \tilde{N}$ are homeomorphic to S^{n-1}. Since the map $\tilde{f}$ is a quasi-isometry and the definition of the boundary at infinity involves geodesic rays, it is natural to wonder how images of geodesic under quasi-isometries look like in hyperbolic space. The image of a geodesic under a (K, ε)-quasi-isometry is called a (K, ε)***-quasi-geodesic***. We add the constraint that quasi-geodesics be continuous. We do so because they will always arise as images of geodesics under the quasi-isometry $\tilde{f}$ which is continuous.

Quasi-geodesics in Euclidean space and hyperbolic space behave differently. It is easy to check that the logarithmic spiral in $\mathbb{R}^2$ is a quasi-geodesic that moves arbitrarily far from straight lines. The next lemma shows that this cannot happen in hyperbolic space.

Lemma 2.8. Morse-Mostow Lemma. *Given $K > 0$, there exists a constant $D = D(K) > 0$ such that for any (K, ε)-quasi-geodesic $\beta : \mathbb{R} \to \mathbb{H}^n$, there exists a unique geodesic $\gamma : \mathbb{R} \to \mathbb{H}^n$ within a D-neighborhood of β.*

Proof. We first prove a local statement. Let $I = [a, b]$, $x = \beta(a)$, $y = \beta(b)$ and γ be the geodesic segment from x to y. Pick $D \gg K$ and suppose that β does not stay at distance D from γ. Let x', $y' \in \beta(I)$ be distinct points at distance D from γ, as shown on Figure 2.2. Let β' be the segment from x' to y'. An elementary calculation shows that

$$l(\beta') \leq K^2 d(x', y') + \varepsilon K.$$

This computation uses the fact that our quasi-geodesics are continuous. Now let γ' be the geodesic segment joining the projection of x' and y' on γ. It is a classical fact that projections in hyperbolic space decrease distances exponentially (see [20, Lemma 11.8.4]). This together with the triangle inequality implies that

$$\begin{aligned} l(\beta') &\leq K^2(l(\gamma') + 2D) + \varepsilon K \\ &\leq K^2(e^{-D} d(x', y') + 2D) + \varepsilon K. \end{aligned}$$

Notice that $d(x', y') \leq l(\beta')$ and use that $D \gg K$ to obtain

$$l(\beta') \leq \frac{2DK^2 + \varepsilon K}{1 - K^2 e^{-D}} \leq 4D^2.$$

Thus β stays in a $D + 4D^2$ neighborhood of γ. Note that this estimate depends only on K, so this holds for any bounded interval.

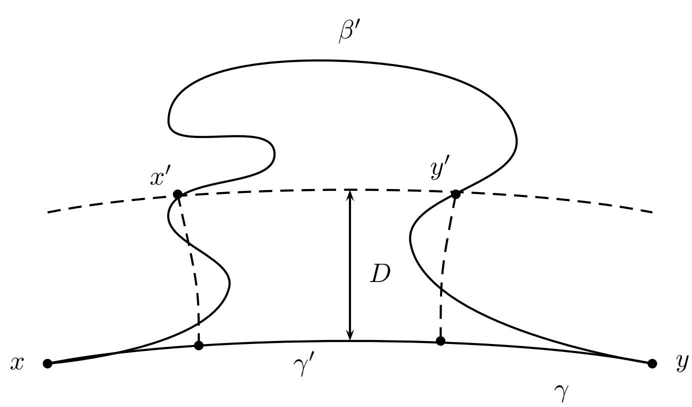


Figure 2.2: Proof of Morse-Mostow Lemma

The proof of the global existence of γ is essentially an application of Arzela-Ascoli theorem. The idea is to define $\gamma_m : \mathbb{R} \to \mathbb{H}^n$ to be the geodesic line through $\beta(-m)$ and $\beta(m)$, where $m \in \mathbb{N}$ (parametrize it carefully so that the points $\gamma_m(0)$ does not move too much). Then restrict to a big compact set and apply the compactness theorem to find a subsequence converging uniformly to some geodesic γ that can be extended to $\mathbb{R}$. The details are left to the reader.

Uniqueness comes from the fact that distinct geodesics in hyperbolic space do not stay at bounded distance from each other. □

We can now define the boundary map $\partial\tilde{f} : S^{n-1} \to S^{n-1}$. For a geodesic γ in $\tilde{M}$, let γ' be the unique geodesic in $\tilde{N}$ close to $\tilde{f}(\gamma)$. Now let

$$\partial\tilde{f}(\gamma(\infty)) = \gamma'(\infty).$$

To check that $\partial\tilde{f}$ is well-defined, observe that if geodesics γ, τ are asymptotic, then the unique geodesics close to $f(\gamma)$ and $f(\tau)$ are also asymptotic. A simple but crucial observation is that the map $\partial\tilde{f}$ is still $\pi_1(M)$-equivariant. We will use the same symbol for the induced action of $\pi_1(M)$ on $\partial\tilde{M}$. Thus equivariance may be rewritten as

$$\partial\tilde{f}(\gamma.\xi) = \rho(\gamma).\partial\tilde{f}(\xi) \text{ for all } \xi \in \partial\tilde{M}.$$

The map $\partial\tilde{f}$ is one-to-one. To see this, pick $\xi \neq \eta \in \partial\tilde{M}$, let γ be the geodesic line such that $\gamma(-\infty) = \xi$ and $\gamma(\infty) = \eta$. Since $\partial\tilde{f}(\xi)$ and $\partial\tilde{f}(\eta)$ correspond to endpoints of some geodesic line, they must be distinct. Besides, since $\tilde{f}$ is a quasi-isometry, it is coarse onto and so every point in $\partial\tilde{N}$ is an accumulation point of $\tilde{f}(\tilde{M})$. This proves surjectivity.

The continuity of $\partial\tilde{f}$ follows from a geometrical lemma that again uses negative curvature and the fact that any two points at infinity can be joined by a geodesic (in other words, the fact that $\mathbb{H}^n$ is a visibility space).

Lemma 2.9. *Let γ be a geodesic in $\mathbb{H}^n$ and P some hyperplane orthogonal to γ. Let γ' be the geodesic close to $\tilde{f}(\gamma)$. Then there exists a constant $D > 0$ depending only on the quasi-isometry constants of $\tilde{f}$ such that*

$$\mathrm{Diam}(\mathrm{Proj}_{\gamma'}(\tilde{f}(P))) \leq D.$$

Notice that this is false in a Euclidean space. A quasi-isometry can tilt γ and P so that the projection of $\tilde{f}(P)$ has infinite diameter.

Proof. Let $x_0 = \gamma \cap P$ and let β be a geodesic in P passing through x_0. Let τ be the geodesic joining $\beta(\infty)$ and $\gamma(\infty)$. There is an absolute constant $A > 0$ such that $d(x_0, \tau) \leq A$. Let γ', β', τ' be the geodesics close to $\tilde{f}(\gamma)$, $\tilde{f}(\beta)$, $\tilde{f}(\tau)$ respectively. Let x_0' be the point on γ' closest to $\tilde{f}(x_0)$. Let $\beta^\perp$ be the geodesic orthogonal to γ' with endpoint $\beta'(\infty)$ and let $y_0 = \gamma' \cap \beta^\perp$.

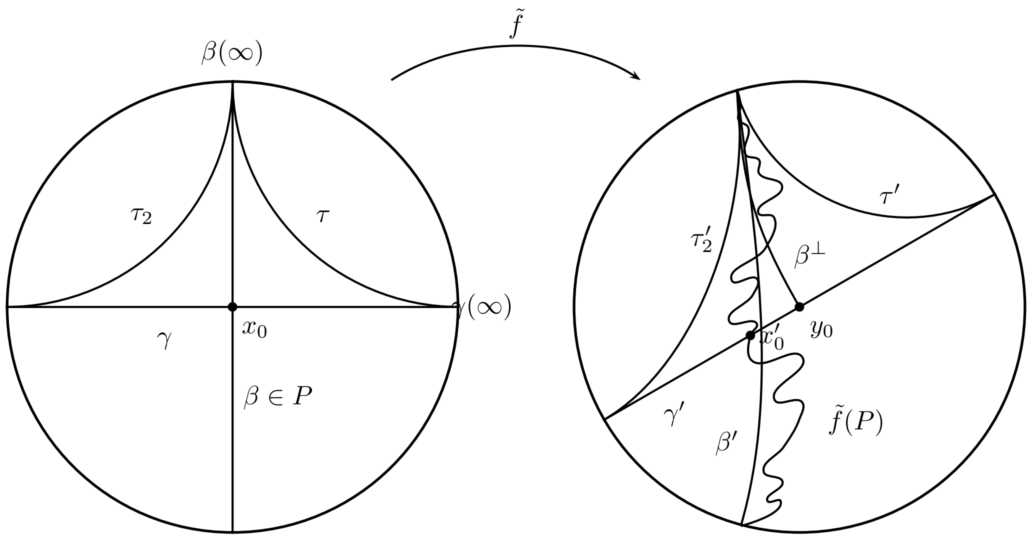


Figure 2.3: The diameter of the projection of an hyperplane

Since $\tilde{f}$ is a (K, ε)-quasi-isometry, we have $d(\tilde{f}(x_0), \tilde{f}(\tau)) \leq KA + \varepsilon$. By Morse-Mostow Lemma 2.8, we have $d(x_0', \tilde{f}(x_0)) \leq L$ and $d(\tau', \tilde{f}(\tau)) \leq L$ for some constant $L > 0$, so that $d(x_0', \tau) \leq KA + \varepsilon + 2L =: D'$. Therefore, on one side of x_0' (the left side on Figure 2.2), the point y_0 must be at distance at most D' from x_0'. Applying the same argument to the geodesic τ_2 joining $\gamma(-\infty)$ and $\beta(\infty)$, we infer that $d(x_0', y_0) \leq D'$. Therefore, the projection of β' onto γ' lies within distance D' from x_0'. Since any $y \in \tilde{f}(\beta)$ lies at distance L from β' and since orthogonal projections decrease distances, it follows that $d(x_0', \mathrm{Proj}_{\gamma'}(y)) \leq D' + L$. Since β can be arbitrarily chosen in P, the lemma follows.

□

Proposition 2.10. *The boundary map $\partial \tilde{f} : \mathbb{H}^n \to \mathbb{H}^n$ is a homeomorphism.*

Proof. For a geodesic ray $\gamma \in \mathbb{H}^n$, let $\chi = \gamma(\infty)$. In the ball model, a basis of neighborhoods of $\partial\tilde{f}(\chi)$ in $\mathbb{H}^n$ (in the cone topology) is given by intersections between the ball and the interior of some hyperbolic hyperplane Q orthogonal to γ', where γ' is as usual the geodesic close to $\tilde{f}(\gamma)$.

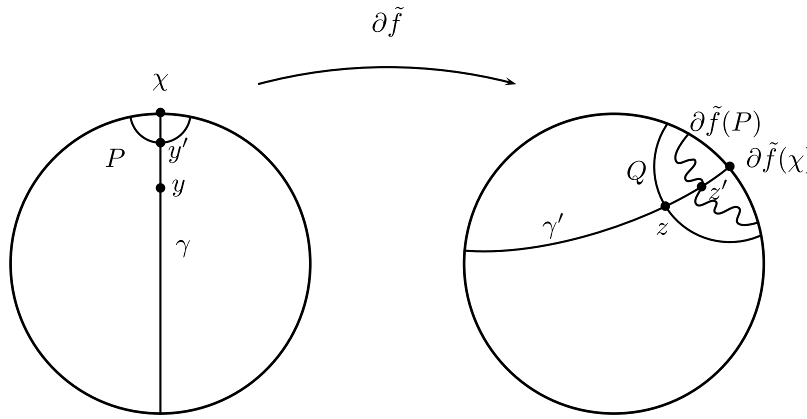


Figure 2.4: The boundary map is a homeomorphism

Our goal is to find a hyperplane $P \perp \gamma$ such that $\tilde{f}(P)$ is contained in the half-space delimited by Q and containing $\partial\tilde{f}(\chi)$. Let $z = Q \cap \gamma$. Let $y \in \gamma$ such that $\tilde{f}(y)$ lies at distance at most L from z, where L is the constant arising in Morse-Mostow Lemma 2.8. Morally it suffices to choose a hyperplane P orthogonal to γ such that the point $y' := P \cap \gamma$ lies sufficiently far away from y. Let z' be the projection of $\tilde{f}(y')$ onto γ'. By Lemma 2.9 it is enough to ensure that $d(z, z') \geq D$ (see Figure 2.4). This can be done easily since quasi-isometries provoke only bounded perturbation. The interested reader can work out the constant that works.

Finally, this argument can be reversed to show that the inverse of $\partial\tilde{f}$ is continuous. □

2.3 Properties of the boundary map

The properties of $\partial\tilde{f}$ shown in this section will only be used in Thurston's proof of the theorem. The two other proofs will just rely on the homeomorphism property. Out of these three proofs, this one is the most analytical, since it relies on a hard analysis fact and on some ergodic theory.

For simplicity, we are going to write $h = \partial\tilde{f}$ throughout.

Quasi-conformality

Our ultimate goal in this section is to show that h is a conformal map. The proof of this will not be direct. Conformality will be a consequence of the (a.e.) differentiability of h and some ergodic theory. Differentiability will follow from a weaker property, namely quasi-conformality.

Definition 2.11. Let X,Y be metric spaces and $K > 0$. A map $f : X \to Y$ is K***-quasi-conformal*** if for all $x \in X$,

$$\lim_{r\to 0} \frac{\sup_{z\in X:d(x,z)=r} d(f(x), f(z))}{\inf_{z\in X:d(x,z)=r} d(f(x), f(z))} < K.$$

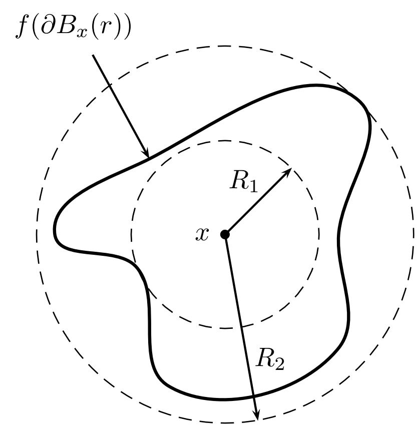


Figure 2.5: Definition of quasi-conformality

With the schematic of Figure 2.3, the function f is quasi-conformal if $\lim_{r\to 0} \frac{R_2}{R_1} < K$ for some $K > 0$. It is easy to see that the composition of two quasi-conformal maps is again quasi-conformal and that the constants multiply. In particular, composing a K-quasi-conformal map with a conformal map again yields a K-quasi-conformal map.

Proposition 2.12. *The boundary map is quasi-conformal.*

Proof. We work out this proof in the upper-half space model which is conformally equivalent to the ball model. Let $\xi \in S^{n-1}$. Up to an isometry, we assume that ξ is at the origin. Let l be the vertical line through ξ and let P be a small Euclidean half-sphere around ξ. It defines a neighborhood of ξ in S^{n-1}. We will show that this neighborhood cannot be deformed too much by h. Composing h with an isometry, we can assume that it fixes ξ and l. By Lemma 2.9, the projection of $h(P)$ onto l has bounded diameter D, where D only depends on the quasi-isometry constants. Let Q and Q' be Euclidean $(n-1)$-half spheres that bound the projection of $\partial\tilde{f}(P)$.

At this stage, one should be careful about which distance is considered. We are studying the map h, which is defined on S^{n-1} endowed with the usual spherical metric. The Euclidean metric on the horizontal hyperplane $\{x_n = 0\}$ obviously does not correspond to the spherical metric. However, this does not matter since these metrics are close to each other in a small neighborhood of ξ. Thus let R_1, resp. R_2, be the radii in the Euclidean metric of Q, resp. Q'. Then

$$D \geq \int_{R_1}^{R_2} \frac{dx}{x} = \log(R_2/R_1),$$

so that h is an e^D-quasi-conformal map. □

The following hard analysis fact will be essential to us. It follows from a result in [21].

Theorem 2.13. *Suppose $n \geq 3$. A quasi-conformal homeomorphism $h : S^{n-1} \to S^{n-1}$ is differentiable almost everywhere. Moreover, the differential is uniformly bounded a.e., that is, there exists a constant $\lambda > 1$ such that for a.a. $x \in S^{n-1}$ and for all $v \in T^1_x S^{n-1}$,*

$$\frac{1}{\lambda} \leq \frac{\|D_x h(v)\|}{\|v\|} \leq \lambda.$$

Notice that this is the only part of the proof that uses $n \geq 3$. It is interesting to see where the argument fails when $n = 2$. In this case, we consider homeomorphisms of S^1. It is a classical fact that any such map is differentiable a.e.. However, the condition on the differential need not hold. Namely, there exist homeomorphisms of the circle that have zero derivative at each point of differentiability. This very fact prevents from using the next argument.

We now turn back to the case $n \geq 3$. For a.a. $x \in S^{n-1}$, consider the tangent space $T_x S^{n-1}$. Since the differential of h at x is uniformly bounded, the image of the unit sphere in $T_x S^{n-1}$ maps to a nondegenerate ellipsoid in $T_{h(x)} S^{n-1}$. Let $v_1(h(x)), \dots, v_{n-1}(h(x))$ be the principal vectors of this ellipsoid. Normalize these vectors so that $\prod_{i=1}^{n-1} \|v_i(h(x))\| = 1$.

Let us now define the excentricity function

$$e_h(x) = \max_{i \neq j} \left\{ \frac{\|v_i(h(x))\|}{\|v_j(h(x))\|} \right\} \quad \text{for a.a. } x \in S^{n-1}.$$

Now notice that $e_h(\gamma.x) = e_h(x)$ for all $\gamma \in \pi_1(M)$. To see this, observe that $\pi_1(M)$ acts by conformal maps on the boundary and that the maps $h \circ \gamma$ and $\rho(\gamma) \circ h$ are equal, so that they have the same differential. This means that $e_h(\gamma.x) = e_{\rho(\gamma)\circ h}(x) = e_h(x)$.

We now quote a result that we are going to prove in the next section.

Theorem 2.14. *The diagonal action of $\pi_1(M)$ on $S^{n-1} \times S^{n-1}$ is ergodic. In particular, $\pi_1(M)$ acts ergodically on S^{n-1}.*

The reader who is new to ergodicity should believe for the moment that this implies that $e_h = c$ a.e. for some $c \geq 1$. It is an easy consequence of ergodicity that will be explained in the next section. We are going to show in Proposition 2.16 that $c = 1$. From the definition of the excentricity function e_h, this means that the differential of h is a conformal map. We need to quote one more analytical result which is addressed in [5].

Theorem 2.15. *If a map $h : S^{n-1} \to S^{n-1}$ is quasi-conformal and its differential Dh is conformal, then h is conformal.*

Accepting these facts for the moment, we can finish the proof of Mostow's rigidity theorem.

Proof of Mostow rigidity. Recalling Theorem 1.4, the fact that the boundary map $h = \partial \tilde{f}$ is conformal means it is induced by some isometry $\tilde{F} : \tilde{M} \to \tilde{N}$. Since isometries of $\mathbb{H}^n$ are uniquely determined by their effect on the boundary, the map $\tilde{F}$ must be $\pi_1(M)$-equivariant. Moreover, $\tilde{F}$ and $\tilde{f}$ induce the same map on the fundamental group of M and so they are homotopic. Thus $\tilde{F}$ descends to an isometry $F : M \to N$ that is homotopic to f and induces the same isomorphism of $\pi_1(M)$. □

We now prove that Dh is conformal, which amounts to prove that $e_h = 1$ a.e..

Proposition 2.16. *The excentricity function e_h equals 1 a.e..*

This will follow from the next result.

Lemma 2.17. *There is no $\pi_1(M)$-invariant measurable frame field defined almost everywhere on S^{n-1}.*

Proof of Proposition 2.16. Assume by contradiction that $e_h = c > 1$ a.e.. Thus for a.a. $x \in S^{n-1}$, there is a measurable frame field $(v_1(x), \dots v_{n-1}(x))$ in $T_x S^{n-1}$, normalized as above. Since h satisfies $h \circ \gamma = \rho(\gamma) \circ h$, we must have

$$\{v_i(\gamma.x)\}_{i=1}^{n-1} = \{D\gamma(v_i(x))\}_{i=1}^{n-1}.[1]$$

Assume now for simplicity that the vectors $(v_1(x), \dots, v_{n-1}(x))$ have different norms for a.e. $x \in S^{n-1}$ and are labeled so that $\|v_1(x)\| < \dots < \|v_{n-1}(x)\|$. As a result, the vectors $(v_1(x), \dots, v_{n-1}(x))$ form a $\pi_1(M)$-invariant measurable frame field on S_{n-1}. But by Lemma 2.17, this is impossible. Hence $c = 1$.

□

Proof of Lemma 2.17. Assume for contradiction that there exists a $\pi_1(M)$-invariant measurable frame field $\{v_1(x), \dots v_{n-1}(x)\}$ for a.e. $x \in S^{n-1}$. For $x, y \in S^{n-1}$, $x \neq -y$, let $P_{yx} : T_y S^{n-1} \to T_x S^{n-1}$ be the parallel translation map along the unique geodesic joining y to x.

For $1 \leq i, j \leq n$, define functions

$$\begin{array}{rcl} \varphi_{i,j} : T_p S^{n-1} \times T_p S^{n-1} & \longrightarrow & \mathbb{R} \\ (x, y) & \longmapsto & \langle v_i(x), P_{yx}(v_j(y)) \rangle. \end{array}$$

These functions are defined a.e.. Since $\pi_1(M)$ acts ergodically on $S^{n-1} \times S^{n-1}$, it follows that the functions $\varphi_{i,j}$ are constant a.e.. Now fix a point $x_0 \in S^{n-1}$ where the frame field is defined. Since the functions $\varphi_{i,j}$ are constant, the vectors $v_1(x_0), \dots, v_n(x_0)$ uniquely determine $v_1(y), \dots, v_n(y)$ for a.e. $y \in S^{n-1}$ via the map S_{xy}. We will now show with an elementary argument that this is impossible.

The proof is based on projections on an isometrically embedded 2-sphere and uses the Gauss-Bonnet theorem. Isometrically embed a 2-sphere S in S^{n-1} such that it contains at least three non-aligned points where the frame field is defined. Denote by $\mathrm{Proj}_{T_x S} : T_x S^{n-1} \to T_x S$ the projection map. It is

[1] Note that the measurability of the frame field uses the assumption that $c > 1$.

easy to see that the projections commute with the parallel translation map. In other words,

$$S_{xy} \circ \mathrm{Proj}_{T_xS} = \mathrm{Proj}_{T_yS} \circ S_{xy} \quad \text{for all } x, y \in S, x \neq -y.$$

Now fix non aligned points $x, y, z \in S$ where the frame field is defined and pick $j \in \{1, \dots, n\}$ such that $\mathrm{Proj}_{T_xS}(v_j(x)) \neq 0$. The invariance property of $v_j(x)$ together with the above commutation property imply that

$$\begin{aligned} \mathrm{Proj}_{T_xS}(v_j(x)) &= \mathrm{Proj}_{T_xS}(S_{zx} \circ S_{yz} \circ S_{xy}(v_j(x))) \\ &= S_{zx} \circ S_{yz} \circ S_{xy}(\mathrm{Proj}_{T_xS}(v_j(x))). \end{aligned}$$

Since the points x, y, z form a non-degenerate spherical triangle, the Gauss-Bonnet theorem tells us that this is impossible. □

Remark 2.18. Lemma 2.17 can be proved using the fact that the restricted holonomy group of a connection ∇ is trivial if and only if ∇ is flat.

2.4 Ergodicity of geodesic flow

We postponed the proof of Theorem 2.14 to this section. It states that the action of $\pi_1(M)$ on S^{n-1} is ergodic. To prove this, we need to introduce some basic facts about ergodicity, Birkhoff's Ergodic Theorem and geodesic and horocyclic flows. Then we show the ergodicity of the geodesic flow on a complete hyperbolic manifold of finite area, from which Theorem 2.14 will follow.

Basics of ergodicity

Definition 2.19. Let (X, μ) be a measure space. A measurable map $T : X \to X$ is said to be ***measure-preserving*** if $\mu(T^{-1}(E)) = \mu(E)$ for all measurable $E \subset X$. A measure-preserving transformation $T : X \to X$ is ***ergodic*** with respect to μ if the only T-invariant subsets have zero or full measure. In other words, T is ergodic if

$$T(E) = E \quad \Rightarrow \quad \mu(E) = 0 \text{ or } \mu(X - E) = 0 \text{ for all measurable } E \subset X.$$

Ergodicity has an equivalent formulation as follows.

Proposition 2.20. *Let X be a finite-measure space. A measure-preserving map $T : X \to X$ is ergodic if and only if every T-invariant function $f \in L^2(X)$ is constant.*

Recall that in the last section, the excentricity function e_h was seen to be $\pi_1(M)$-invariant. Provided the action of $\pi_1(M)$ on S^{n-1} is ergodic (see Theorem 2.14), Proposition 2.20 implies that e_h is a constant function.

A rather short proof of the next theorem can be found in [12].

Theorem 2.21. Birkhoff's ergodic theorem. *Let (X, μ) be a probability space and suppose $T : X \to X$ is a measure-preserving transformation. Then the limit*

$$f^*(x) = \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} f(T^{(k)}(x))$$

exists for a.a. $x \in X$. The function f^* is T-invariant and satisfies

$$\int_X f^* \, d\mu = \int_X f \, d\mu.$$

If T is ergodic, then $f^* = c$ a.e. for some constant $c \in \mathbb{R}$ and

$$c = \int_X f^* \, d\mu = \int_X f \, d\mu.$$

Thus

$$\lim_{N\to\infty} \sum_{k=1}^{N} f(T^{(k)}(x)) = \int_X f \, d\mu \text{ for a.a. } x \in X.$$

Remark 2.22. The last equation is often rephrased as follows. If T is ergodic, then the space average equals the time average. Also note that the conclusions of Birkhoff's ergodic theorem also hold when considering a 1-parameter family of measure-preserving maps $T_s : X \to X$, $s \in \mathbb{R}$. Since we are going to deal with the geodesic flow, this is will be our case of interest.

Geodesic and horocyclic flows

Definition 2.23. Let X be a complete Riemannian n-manifold and denote by T^1X the unit tangent bundle. For $v \in T^1X$, write γ_v for the infinite geodesic with initial tangent vector v. The ***geodesic flow*** on T^1X is the 1-parameter family of diffeomorphisms defined by

$$\begin{aligned} g_t : T^1X &\longrightarrow T^1X \\ v &\longmapsto (\gamma_v(t), D(\gamma_v)_t(v)). \end{aligned}$$

Definition 2.24. Let X be a Riemannian manifold of finite measure. The ***Liouville measure*** on T^1X is defined to be

$$d\omega = d\,\mathrm{Vol}_X\, d\theta,$$

where the volume form of X has unit total mass and $d\theta$ is the Lebesgue measure of unit mass on the $(n-1)$-sphere T^1_pX, for all $p \in X$.

Thus the Liouville measure is a probability measure on T^1X. The geodesic flow preserves the Liouville measure. To see this, first check this on subsets of T^1X of the form T^1U, where U is an open set in X, and observe that g_t preserves the measure of subsets of T^1_pX for all $p \in X$.

From now on, let M be a finite-measure complete hyperbolic n-manifold. For simplicity, we are going to prove the ergodicity of the geodesic flow for $n = 2$. Although this may seem paradoxical (since Mostow rigidity does not hold in this case), the proof easily extends to higher dimensions.

Apart from the geodesic flow g_t, the proof will use two other flows on T^1M. Given a tangent vector $v \in T^1_pM$, consider the geodesic γ with initial tangent vector v. Let $HS^+(v)$ be the horosphere through p centered at $\gamma(\infty)$ and $HS^-(v)$ be the horosphere through p center at $\gamma(-\infty)$. The ***positive horocyclic flow*** h^*_s moves v along $HS^+(v)$ at distance s from p so that $h^*_s(v)$ is a unit vector orthogonal to the positive horosphere. The ***negative horocyclic flow*** is defined similarly, namely $h_u(v)$ is the unit vector orthogonal to $HS^-(v)$ at distance u from p.

This construction is most easily visualized in the upper-half space model, as shown on Figure 2.6. Up to an isometry sending the vector v to i_i, the following identities are easily seen.

$$g_t h^*_s = h^*_{se^{-t}} g_t, \qquad g_t h_s = h_{se^t} g_t. \tag{2.1}$$

Our proof of the ergodicity of the geodesic flow is based on [1].

Theorem 2.25. *Let M be a complete connected hyperbolic manifold of finite measure. Then the geodesic flow on T^1M is ergodic with respect to Liouville measure.*

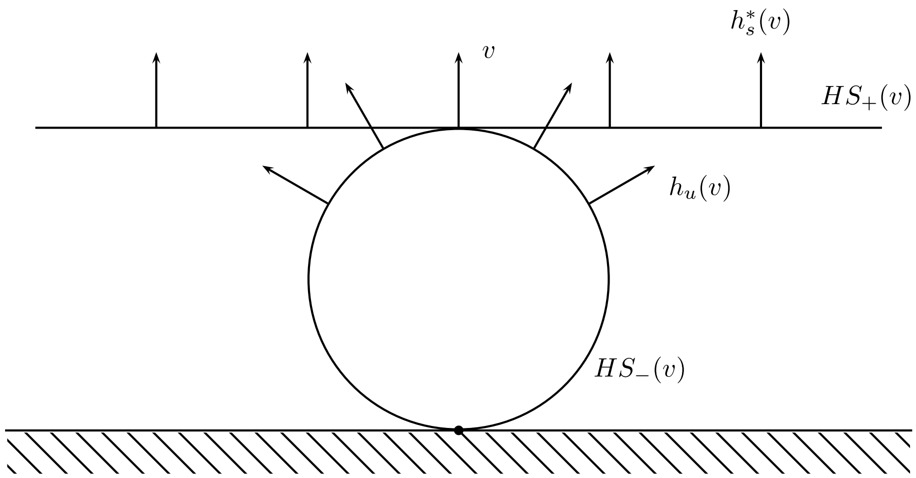


Figure 2.6: Positive and negative horocyclic flows

Proof. The key tool will be a kind of converse to Birkhoff's ergodic theorem. Let f be any function in $L^1(T^1M)$. Since g_t is measure-preserving, Birkhoff's Ergodic Theorem 2.21 implies that for a.a. $v \in T^1M$, the limits

$$\begin{aligned} f^+(v) &= \lim_{T\to\infty} \frac{1}{T}\int_0^T f(g_t(v))\,dt, \\ f^-(v) &= \lim_{T\to\infty} \frac{1}{T}\int_{-T}^0 f(g_t(v))\,dt \end{aligned}$$

exist and $f^+, f^- \in L^1(T^1M)$. Moreover, for a given $v \in T^1M$, whenever either limit exists, we have $f^+(v) = f^-(v)$ (see [13, Corollary II.1.4].

We claim that for any $f \in L^1(T^1M)$, the function f^+ is constant a.e.. This implies the ergodicity of the geodesic flow g_t (see [24]). By a classical density argument, it is enough to prove the claim when f is continuous function with compact support.

The first step is to prove that f^+ (resp. f^-) is constant along positive (resp. negative) horocycles. Since f is uniformly continuous, there is a $\delta > 0$ such that $|f(v_1) - f(v_2)| < \varepsilon$ whenever v_1 and v_2 are δ-close. Pick a $v \in T^1M$ such that $f^+(v)$ exists. Fix $s \in \mathbb{R}$ and let $t_0 \in \mathbb{R}$ such that $|s|e^{-t_0} < \delta$. Then (2.1) implies that

$$d(g_t(v), g_t h_s^*(v)) = d(g_t(v), h^*_{se^{-t}} g_t(v)) = |s|e^{-t} < \delta \text{ for all } t \geq t_0.$$

Therefore

$$\frac{1}{T}\int_0^T f(g_{t+t_0}(v))\,dt \quad \text{and} \quad \frac{1}{T}\int_0^T f(g_{t+t_0}(h_s^*(v)))\,dt$$

are ε-close. Since ε is arbitrary, this means that $f^+(v) = f^+(h_s^*(v))$ for all $s \in \mathbb{R}$. The proof for f^- is exactly the same.

Now define $W^s(v) = \{g_t h_s^*(v) : s, t \in \mathbb{R}\}$ and $W^u(v) = \{g_t h_s(v) : s, t \in \mathbb{R}\}$. The sets $W^s(v)$ and $W^u(v)$ are called the stable (resp. unstable) foliations. The last paragraph together with the fact that $f^+(g_t(v)) = f^+(v)$ a.e. imply that f^+ (resp. f^-) is constant on $W^s(v)$ (resp. $W^u(v)$).

At this point, for distinct v_1, $v_2 \in T^1M$, the function f^+ might assume different values on $W^u(v_1)$ and $W^u(v_2)$. But the fact that one can almost always travel from $W^u(v_1)$ to v_2 along a positive horocycle will imply that this cannot happen. More precisely, we claim that

$$\omega(\{h_s^* g_t h_a(v) : a, s, t \in \mathbb{R}\}) = 1. \tag{2.2}$$

(and we know exactly the set where it does not hold).

To see this, let w_1, w_2 be in T^1M and consider the unique geodesic γ that cuts orthogonally $HS^-(w_1)$ and $HS^+(w_2)$. Figure 2.7 shows that we can travel from w_1 to w_2 by moving successively along the negative horocyclic flow, the geodesic flow and the positive horocyclic flow. The picture is similar in higher dimensions (horosphere will simply have dimension $n-1$).

The attentive reader will have noticed that this construction fails if and only if $w = -v'$, for some element $v' \in W^s(v)$. But this subset of T^1M has zero measure.

Similarly,

$$\omega(\{h_a g_t h_s^*(v) : a, s, t \in \mathbb{R}\}) = 1. \tag{2.3}$$

Observe that around each point w in T^1M, there is a neighborhood diffeomorphic to a small neighborhood of the origin in $\mathbb{R}^3$ via the map

$$(s, t, u) \mapsto h_s^* g_t h_u(w).$$

In a neighborhood of w, these coordinates provide us with a measure $dsdtdu$ that is equivalent to the Liouville measure $d\omega$, in the sense of having the same sets of measure zero. This can be seen using a figure similar to Figure 2.7. As a result, there exists a full Lebesgue measure subset U of $\mathbb{R}$ such that $f^+(h_u(v))$ and $f^-(h_u^*(v))$ exist for all $u \in U$.

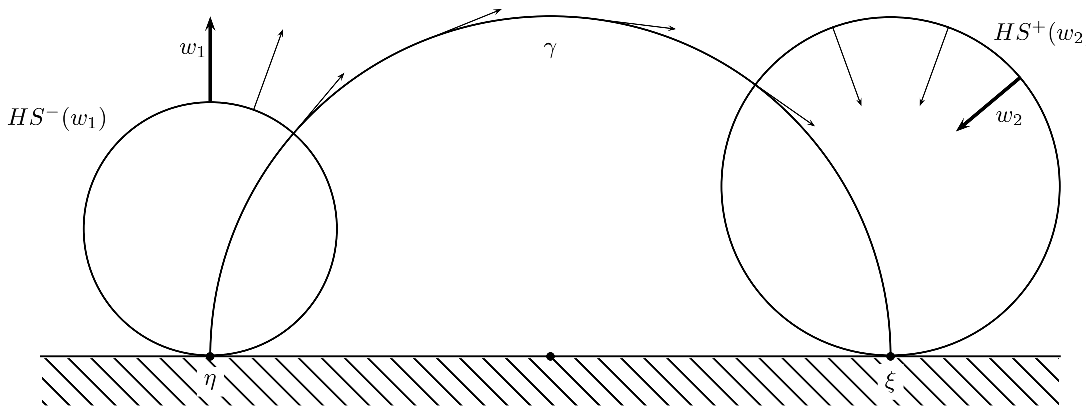


Figure 2.7: How to travel on horocyclic and geodesic flows

All these partial results sum up as follows. For all $u \in U$,

$$f^+ \text{ is constant on } W^s(h_u(v)) \text{ and exists on } W^u(h_u^*(v)), \tag{2.4}$$

$$f^- \text{ is constant on } W^u(h_u^*(v)) \text{ and exists on } W^s(h_u(v)) \tag{2.5}$$

Finally, for u_1, $u_2 \in U$,

$$\begin{aligned} f^+(h_{u_2}(v)) &= f^-(h_{u_2}(v)) && \text{since } u_2 \in U \text{ and } f^+,\ f^- \text{ agree whenever they exist} \\ &= f^-(h_{u_1}(v)) && \text{since } f^- \text{ is constant along negative horocycles} \\ &= f^+(h_{u_1}(v)) && \text{since } u_1 \in U \text{ and } f^+,\ f^- \text{ agree whenever they exist,} \end{aligned}$$

and so (2.2) implies that f^+ is constant a.e.. □

Now we again assume M to be a *compact* hyperbolic manifold of dimension $n \geq 2$. Everything is ready to prove Theorem 2.14 which states that the diagonal action of $\pi_1(M)$ on $S^{n-1} \times S^{n-1}$ is ergodic.

Proof of Theorem 2.14. We want to translate our knowledge about the ergodicity of the geodesic flow on M to the boundary of $\tilde{M} = \mathbb{H}^n$. This can be done by putting certain coordinates on $T^1\mathbb{H}^n$. For a unit vector v tangent to $p \in \mathbb{H}^n$, let φ be the geodesic with initial tangent vector v, let ξ, η be its endpoints and define $s \in \mathbb{R}$ to be the hyperbolic signed distance between p and the Euclidean midpoint of φ. The coordinates that we will use are

$$\begin{array}{rcl} T^1\mathbb{H}^n & \longrightarrow & (S^{n-1} \times S^{n-1} - \Delta) \times \mathbb{R} \\ v & \longmapsto & (\xi, \eta, s). \end{array}$$

We need to know how the Liouville measure on $\mathbb{H}^n$ behaves under this change of coordinates. It is a fact that there exists a positive function $\rho(\xi, \eta)$ such that

$$d\omega = \rho(\xi, \eta) d\xi d\eta ds.$$

Let $A \subset S^{n-1} \times S^{n-1} - \Delta$ be a $\pi_1(M)$-invariant subset. Since the diagonal Δ has zero measure, one can forget about it. Define $B = A \times \mathbb{R}$ and observe the following.

1. B is invariant under the geodesic flow on $\mathbb{H}^n$,
2. B is $\pi_1(M)$-invariant (since A is $\pi_1(M)$-invariant),
3. the geodesic flow on $M = \mathbb{H}^n/\pi_1(M)$ is the projection of the geodesic flow on $\mathbb{H}^n$.

The last fact holds because for all $\gamma \in \pi_1(M)$ and all geodesic ψ in $\mathbb{H}^n$, the commutation property

$$\psi_t \circ D\gamma = D\gamma \circ \psi_t$$

holds. As a result, the set B/Γ is invariant under the geodesic flow on M. By Theorem 2.25, either $\omega(B/\Gamma) = 0$ or $\omega(M - B/\Gamma) = 0$. In the former case, we get

$$0 = \omega_{\mathbb{H}^n}(B) = \int_B d\omega = \int_B \rho(\xi, \eta) d\xi d\eta ds.$$

Therefore,

$$\int_A \rho(\xi, \eta) d\xi d\eta = 0$$

so that $\mu(A) = 0$. The latter case is handled similarly. $\square$

Chapter 3

Gromov's proof

As pointed out above, the only feature that Gromov's proof of Mostow rigidity shares with Thurston's proof is the use of the boundary map. In this chapter, we introduce a homological invariant of a manifold known as Gromov's norm. Gromov's norm of hyperbolic manifolds will be seen to be proportional to the volume of the manifold. The first striking consequence of this result is that the volume of a hyperbolic manifold is a topological invariant. To deduce Mostow rigidity, one has to show that the boundary map preserves the family of ideal simplices of maximal volume.

Simplices in hyperbolic space and a modified homology where cycles are measures are the major tools in this chapter. We follow Thurston [22] and Munkholm [17].

3.1 Gromov norm

We start by fixing some notation and introducing the Gromov norm of a manifold. For a topological space X, let $C_*(X)$ be the real singular chain complex. Any k-chain can be written uniquely as

$$c = \sum_i a_i \sigma_i,$$

where $a_i \in \mathbb{R}$ and $\sigma_i : \Delta^k \to X$ is a C^1 map. Here Δ^k denotes the standard k-simplex in $\mathbb{R}^n$.

We endow $C_*(X)$ with the l^1 norm. That is, the norm of a k-chain c as

above is
$$\|c\| = \sum_i |a_i|.$$
This norm descends to a seminorm on the real homology groups $H_k(X)$ by taking the infimum. Namely, for $z \in H_k$,
$$\|z\| = \inf\{\|c\| : c \text{ is a singular } k\text{-chain representing } z\}.$$
The axioms of a seminorm are easily verified. Also notice that the induced action of a continuous map $f : X \to Y$ on homology classes decreases the norm. In other words, for $z \in H_k(X)$,
$$\|f_*(z)\| \leq \|z\|.$$
There is a inequality because if σ_1, σ_2 are distinct simplices in X such that $f \circ \sigma_1 = f \circ \sigma_2$, then
$$\|f_*(\sigma_1 - \sigma_2)\| = 0 < \|\sigma_1 - \sigma_2\|.$$

It is a classical fact from algebraic topology that if M is an orientable n-manifold, then $H_n(M;\mathbb{Z}) = \mathbb{Z}$ (see [11, Theorem 3.26]. A generator for this homology group is called a ***fundamental class*** of M and is denoted by $[M]$. This carries over to real homology by noticing that $H_n(M,\mathbb{R}) = H_n(M,\mathbb{Z}) \otimes \mathbb{R}$.

Definition 3.1. Let M be an orientable manifold. The ***Gromov norm*** of M is defined to be
$$\|M\| = \|[M]\|.$$

The Gromov norm is also called ***simplicial volume***. Note that the existence of manifolds with positive Gromov norm is not obvious. This is emphasized by the next result.

Proposition 3.2. *Suppose $f : M \to N$ is a continuous map between orientable manifolds. Then*
$$\|M\| \geq |\deg f|\|N\|.$$

Proof. Write $d = \deg f$. Since $f_*([M]) = d[N]$, for any cycle z representing $[M]$, the cycle $f_*(z)$ represents $d[N]$. Therefore
$$|d|\|N\| = \|d[N]\| \leq \|f_*(z)\| \leq \|z\|.$$
Conclude by taking the infimum. □

Now suppose that M admits a self-map $f : M \to M$ with $|\deg f| > 1$. The proposition implies that $\|M\| > |\deg f|\|M\|$, so $\|M\|$ must be zero. For example, the n-sphere S^n admits self-maps of any degree (see [11, Ex. 2.31]), so $\|S^n\| = 0$ for all n. However, we are going to see that hyperbolic manifolds have positive Gromov norm.

3.2 Simplices and their volume

Before going into proofs of deeper results, we need to better understand when the infimum in Definition 3.1 is attained. In fact, among all simplices, it is enough to take the infimum over straight simplices.

Suppose that M is a hyperbolic n-manifold. Given a simplex $\sigma : \Delta^k \to M$ that may have a complicated shape, we want to construct a new “nicer” straight simplex. Pick a lift $\tilde{\sigma} : \Delta^k \to \mathbb{H}^n$ of σ and denote its vertices by $v_0, \ldots, v_{k+1}$. Using the hyperboloid model, we can build an affine simplex $\tau : \Delta^k \to \mathbb{R}^{n+1}$ with vertices $v_0, \ldots, v_{k+1}$. The projection of τ onto $\mathbb{H}^n$ is denoted by $\operatorname{str}(\tilde{\sigma})$ and we define $\operatorname{str}(\sigma)$ to be the projection of $\operatorname{str}(\tilde{\sigma})$ onto M. Straightening is extended linearly to $C_*(M)$.

Since isometries of $\mathbb{H}^n$ are affine maps of $\mathbb{R}^{n+1}$, the straightening of σ is independent from the choice of its lift. While dealing with Gromov norm, it is enough to focus on straight simplices. Indeed, for a cycle z,

$$\|\operatorname{str}(z)\| \leq \|z\|.$$

This holds because $\operatorname{str} z$ and z are chain homotopic. Again, there is an inequality since cancellations may occur while straightening.

The maximal volume of simplices will be of special interest. Define

$$v_n = \sup\{\operatorname{Vol}(\sigma) : \sigma \text{ is a straight } n\text{-simplex}\}.$$

Hyperbolic spaces have the property that this quantity is bounded. Asymptotic formulas for v_n are available, but the following estimate will be sufficient for our purposes.

Proposition 3.3. *For $n \geq 2$,*

$$v_n \leq \frac{\pi}{(n-1)!}.$$

Proof. First observe that any straight simplex σ in $\mathbb{H}^n$ is contained in a simplex with vertices at infinity (we now consider maps $\Delta^n \to \overline{\mathbb{H}^n}$). The geometric construction is shown on Figure 3.1. Such a simplex is called an ***ideal simplex***. Thus, in this proof it suffices to consider ideal simplices.

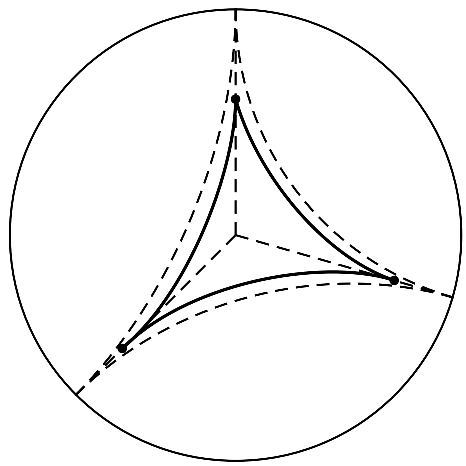


Figure 3.1: Every simplex is contained in an ideal simplex

The proof is by induction. By the Gauss-Bonnet theorem, one has $v_2 = \pi$. The proof will follow from the inequality $v_n \leq \frac{v_{n-1}}{n-1}$.

To see this, work in the upper half-space model and assume that $v_0 = \infty$. Denote by σ_0 the lower $(n-1)$-subsimplex of σ and let τ be the projection of σ onto the horizontal hyperplane. For $z \in \tau$, write $h(z)$ for the Euclidean distance between z and the point above z in σ_0. Thus the volume of σ_0 can be written as

$$\begin{aligned} \mathrm{Vol}(\sigma) &= \int_\tau \int_{h(z)}^\infty \frac{dy}{y^n}\, dz \\ &= \frac{1}{n-1} \int_\tau \frac{1}{h(z)^{n-1}}\, dz \end{aligned}$$

Up to an isometry, we can assume that σ_0 lies in the unit Euclidean upper hemisphere centered at the origin, so that $h(z) = \sqrt{1 - z^2}$. It is now enough to prove that

$$\int_\tau \frac{1}{h(z)^{n-1}}\, dz \leq \mathrm{Vol}(\sigma_0).$$

Let $f : D^{n-1} \to \mathbb{R}^n$ be the parametrization of the unit half sphere given by

$$(x_1, \ldots, x_{n-1}) \mapsto (x_1, \ldots, x_{n-1}, \sqrt{1 - (x_1^2 + \ldots + x_{n-1}^2)}).$$

The volume of σ_0 is given by

$$\mathrm{Vol}(\sigma_0) = \int_\tau \alpha(x) \frac{dx}{h(x)^{n-1}},$$

where $\alpha(x) = \sqrt{\det(\langle Df_x(e_x^i), Df_x(e_x^j)\rangle)}$. An easy computation yields

$$\langle Df_x(e_x^i), Df_x(e_x^j)\rangle = \delta_{ij} + \frac{x_i x_j}{1 - \|x\|^2}.$$

Taking the determinant, we get

$$\alpha^2(x) = 1 + \frac{\|x\|^2}{1 - \|x\|^2} = \frac{1}{h^2(x)}.$$

Therefore, using that $h(x) \leq 1$ for all $x \in \tau$,

$$\mathrm{Vol}(\sigma_0) = \int_\tau \frac{dx}{h(x)^n} \geq \int_\tau \frac{dx}{h(x)^{n-1}}.$$

□

3.3 Gromov norm of hyperbolic manifolds

At the end of Section 3.1, we noticed that the existence of manifolds with nonzero Gromov norm is not obvious. Nevertheless, hyperbolic manifolds have the property that their Gromov norm is bounded away from zero. Moreover, the next theorem we state provides an exact formula to compute Gromov's invariant. This important result is also an essential ingredient in Gromov's proof of Mostow rigidity.

Theorem 3.4. *Let M be a finite-volume hyperbolic n-manifold. Then*

$$\|M\| = \frac{\mathrm{Vol}(M)}{v_n}.$$

This theorem can be proved using the usual singular homology (see [2]). However, the proof is easier to understand when using a modified homology which is a smoothing of singular homology.

Instead of continuous maps from the standard k-simplex, the chains of this modified homology are measures compactly supported on the space $C^1(\Delta^k, M)$. More precisely, a k-chain is a signed Borel measure on $C^1(\Delta^k, M)$ with bounded total variation and compactly supported. Recall that any measure space (X, μ) admits a canonical splitting (X_+, μ_+), (X_-, μ_-) such that $X_+ \cap X_- = \emptyset$ and $\mu = \mu^+ - \mu^-$, where μ^+ and μ^- are nonnegative measures. Then the total variation of μ is given by

$$\|\mu\| = \int_{X_+} d\mu^+ + \int_{X_-} d\mu^-.$$

Let $\mathcal{C}^k(M)$ be the corresponding set of k-chains. The natural face inclusions $\eta_i : \Delta^{k-1} \to \Delta^k$ induce boundary maps in the following way. The map

$$\eta_i^* : C^1(\Delta^k, M) \to C^1(\Delta^{k-1}, M)$$

pushes forward to another map

$$\begin{array}{rcl} \xi_i : \mathcal{C}^k(M) & \longrightarrow & \mathcal{C}^{k-1}(M) \\ \mu & \longmapsto & \xi_i(\mu) = (\eta_i^*)_* \mu. \end{array}$$

The boundary map is defined as $d_k = \sum_{i=0}^k (-1)^i \xi_i$. It is an exercise to show that $d_{k-1} d_k = 0$.

We claim that the natural inclusion $i : C_*(M) \to \mathcal{C}_*(M)$ that sends $\sigma \in C^1(\Delta^k, M)$ to the Dirac measure on σ is a chain map. We are to show that $i(\partial\sigma) = d(i(\sigma))$. To see this, observe that $i(\partial\sigma) = \sum_{j=0}^k (-1)^j \delta_{\sigma_j}$, where δ_{σ_j} stands for the Dirac measure on the j-th side of σ. On the other side, for a Borel subset A of $C^1(\Delta^{k-1}, M)$,

$$\xi_j(i(\sigma))(A) = \left\{ \begin{array}{l} 1 \text{ if } \sigma_j \in A \\ 0 \text{ otherwise} \end{array} \right. = \delta_{\sigma_j}(A)$$

Thus the inclusion map descends to a map between the homology groups. In fact,

Proposition 3.5. *The map*

$$i_* : H_n(C_*(M)) \longrightarrow H_n(\mathcal{C}_*(M))$$

is an isomorphism for all $n \geq 0$.

The de Rham pairing naturally extends to the modified homology by integrating one more time over the set of singular simplices. Denote by $C^*_{DR}(M)$ the de Rham cochain complex. Then the de Rham pairing

$$\begin{aligned} C_k(M) \times C^k_{DR}(M) &\longrightarrow \mathbb{R} \\ (\sigma, \alpha) &\longmapsto \int_M \sigma \, d\alpha. \end{aligned}$$

extends to

$$\begin{aligned} \mathcal{C}_k(M) \times C^k_{DR}(M) &\longrightarrow \mathbb{R} \\ (\mu, \alpha) &\longmapsto \int_{\tau \in C^1(\Delta^k, M)} \left(\int_{\Delta^n} \tau \, d\alpha \right) d\mu(\tau). \end{aligned}$$

Let $p : \mathbb{H}^n \to M$ be a covering map and let Ω_M (resp. $\Omega_{\mathbb{H}^n}$) be the volume form of M (resp. $\mathbb{H}^n$). Also observe that the straightening map extends linearly to a map $C_*(M) \to C_*(M)$ that induces another map $\mathrm{str}_* : \mathcal{C}_*(M) \to \mathcal{C}_*(M)$. Recall that straightening commutes with the projection p.

We are now in position to show that $\mathrm{Vol}(M)/v_n \leq \|M\|$. Let μ be a representative for $[M]$ corresponding to a triangulation of M. Let $\tilde{\tau}$ be a lift of τ to $\mathbb{H}^n$. Then

$$\begin{aligned} \mathrm{Vol}(M) &= \langle \mu, \Omega_M \rangle \\ &= \int_{\tau \in C^1(\Delta^k, M)} \left(\int_{\Delta^n} \tau^* \Omega_M \right) d(\mathrm{str}_* \, \mu) \\ &= \int_{\tau \in C^1(\Delta^k, M)} \left(\int_{\Delta^n} (\underbrace{\mathrm{str}(\tau)}_{= p \circ \mathrm{str} \circ \tilde{\tau}})^* \Omega_M \right) d\mu \\ &= \int_{\tau \in C^1(\Delta^k, M)} \left(\int_{\Delta^n} (\mathrm{str}(\tilde{\tau}))^* \Omega_{\mathbb{H}^n} \right) d\mu \\ &= \int_{\tau \in C^1(\Delta^k, M)} \left(\int_{\mathrm{str}(\tilde{\tau})(\Delta^n)} \Omega_{\mathbb{H}^n} \right) d\mu \\ &\leq v_n \|\mu\|. \end{aligned}$$

Taking the infimum over representatives of $[M]$, one obtains $\mathrm{Vol}(M) \leq v_n \|M\|$.

The reverse inequality requires more work. This is because we are going to explicitly construct a cycle that achieves the bound $\mathrm{Vol}(M)/v_n$. However, having introduced the modified homology on M makes the idea easier to grasp. We will define the smearing operation as follows. Given a singular straight simplex σ in M, we construct a measure $\mathrm{smear}(\sigma)$ that is supported on the set of isometric copies of σ. The cycle that we seek will use the smearing of an ideal simplex of maximal volume.

The following fact will be needed (see [2]).

Proposition 3.6. $\mathrm{Isom}_+(\mathbb{H}^n)$ *is a unimodular Lie group.*

Let h denote a Haar measure on $\mathrm{Isom}_+(\mathbb{H}^n)$. Since $\Gamma := \pi_1(M)$ is a discrete subgroup of $\mathrm{Isom}_+(\mathbb{H}^n)$, it is unimodular. Therefore, by a classical theorem about Haar measures on a quotient, h descends to a Haar measure h_M on the quotient $P(M) := \Gamma \backslash \mathrm{Isom}_+(\mathbb{H}^n)$. Normalize h_M so that $h_M(P(M)) = \mathrm{Vol}(M)$. Let $\sigma \in C^1(\Delta^n, \mathbb{H}^n)$ be fixed. Define

$$\begin{aligned} \varphi_\sigma : P(M) &\longrightarrow C^1(\Delta^n, M) \\ \Gamma g &\longmapsto p \circ g \circ \sigma \end{aligned}$$

and

$$\begin{aligned} \mathrm{smear} : C^1(\Delta^n, \mathbb{H}^n) &\longrightarrow \mathcal{C}_n(M) \\ \sigma &\longmapsto \varphi_{\sigma *}(h_M). \end{aligned}$$

Proposition 3.7. *Let* $\sigma \in C^1(\Delta^n, \mathbb{H}^n)$ *be a straight simplex. Then*

1. $\mathrm{smear}(\sigma^{(i)}) = \xi_i \, \mathrm{smear}(\sigma)$,
2. $\mathrm{smear}(g\sigma) = \mathrm{smear}(\sigma)$ *for all* $g \in \mathrm{Isom}_+(\mathbb{H}^n)$,
3. $\|\mathrm{smear}(\sigma)\| = \mathrm{Vol}(M)$,
4. $\langle \mathrm{smear}(\sigma), \Omega_M \rangle = \mathrm{Vol}(\sigma)\,\mathrm{Vol}(M)$.

Proof. Properties (i) and (ii) are straightforward consequences of definitions. Property (iii) is a consequence of the normalization of h_M and its proof similar to the computation for (iv) that follows. By definition,

$$\begin{aligned} \langle \mathrm{smear}(\sigma), \Omega_M \rangle &= \int_{\tau \in C^1(\Delta^n, M)} \left(\int_{\Delta^n} \tau^* \Omega_M \right) d(\varphi_{\sigma *}(h_M))(\tau) \\ &= \int_{\Gamma g \in P(M)} \left(\int_{\Delta^n} \underbrace{(pg\sigma)^* \Omega_M}_{= \sigma^* \Omega_{\mathbb{H}^n}} \right) dh_M(\Gamma g) \\ &= \mathrm{Vol}(\sigma)\,\mathrm{Vol}(M). \end{aligned}$$

□

It is now time to prove the second inequality of Theorem 3.4. Denote by σ^- the reflection of σ through one of its faces. Let

$$\zeta(\sigma) = \frac{1}{2}(\mathrm{smear}(\sigma) - \mathrm{smear}(\sigma^-)).$$

Since $\mathrm{smear}(\sigma)$ and $\mathrm{smear}(\sigma^-)$ are both nonnegative, Proposition 3.7.(iii) implies that $\|\zeta(\sigma)\| = \mathrm{Vol}(M)$. Since σ and σ^- are not conjugate in $\mathrm{Isom}_+(\mathbb{H}^n)$, the chain $\zeta(\sigma)$ is non trivial. However, the faces of σ and σ^- are conjugate in $\mathrm{Isom}_+(\mathbb{H}^n)$, so that $\zeta(\sigma)$ is a cycle. Remembering that $H_n(M;\mathbb{R}) = \mathbb{R}$, there exists $\lambda \neq 0$ such that $\zeta(\sigma)$ represents $\lambda[M]$. It follows Proposition 3.7.(iv) that

$$\mathrm{Vol}(\sigma)\,\mathrm{Vol}(M) = \langle \zeta(\sigma), \Omega_M \rangle = \lambda \langle [M], \Omega_M \rangle = \lambda\, \mathrm{Vol}(M).$$

Therefore $\zeta(\sigma)$ represents $\mathrm{Vol}(\sigma)[M]$. By definition of Gromov norm, this means that

$$\mathrm{Vol}(M) = \|\zeta(\sigma)\| \geq |\,\mathrm{Vol}\,\sigma| \|M\|.$$

Taking the supremum over all straight simplices, one finally obtains

$$\mathrm{Vol}(M) \geq v_n \|M\|.$$

3.4 Gromov's proof of Mostow rigidity

We again turn our attention to the boundary map $h = \partial \tilde{f} : S^{n-1} \to S^{n-1}$. An application of Theorem 3.4 will show that h maps simplices of maximal volume to other simplices of maximal volume. A result by Haagerup and Munkholm then implies that h maps regular ideal simplices to regular ideal simplices. A geometric argument using repeated reflections of these simplices will conclude Gromov's proof of Mostow's rigidity.

Proposition 3.8. *The boundary map h carries vertices of ideal simplices of maximal volume to vertices spanning an ideal simplex of maximal volume.*

Proof. Let σ be an ideal simplex of maximal volume with vertices $v_0, \dots, v_n$. Assume by contradiction that $\mathrm{Vol}(\mathrm{str}(h(\sigma))) < v_n$. Then there exists $\varepsilon > 0$ and open sets $U_i \subset \mathbb{H}^n$ such that

$$\mathrm{Vol}(h(\mathrm{str}(\sigma(u_0, \dots, u_n)))) < v_n - 2\varepsilon \text{ for all } u_i \in U_i.$$

Choose open subsets $V_i \subset U_i$ with the property that the set

$$A(G) = \{g \in \operatorname{Isom}_+(\mathbb{H}^n) : (v_i \in V_i \Rightarrow gv_i \in U_i \ \forall i)\}$$

has positive measure $m_A > 0$. For any $\delta > 0$, there exists $\sigma_0 = \sigma_0(u_0, \dots, u_n)$ with $u_i \in V_i$ and $\operatorname{Vol}(\sigma_0) > v_n - \delta$.

- if $g \in A(G)$, then $\operatorname{Vol}(\operatorname{str}(\tilde{f}(\sigma_0))) < v_n - 2\varepsilon < \operatorname{Vol}(\sigma_0) - 2\varepsilon + \delta$;
- if $g \notin A(G)$, then $\operatorname{Vol}(\operatorname{str}(\tilde{f}(\sigma_0))) < v_n < \operatorname{Vol}(\sigma_0) + \delta$.

Now integrate on $A(G)$ and its complement to find

$$\begin{aligned}
\langle \operatorname{str} \tilde{f}_*(\operatorname{smear}(\sigma_0)), \Omega_N \rangle &= \int_{\tau \in C^1(\Delta^n, N)} \left(\int_{\Delta^n} \tau^* \Omega_N \right) d(\operatorname{str} \tilde{f}_* \varphi_{\sigma_0 *} h_M) \\
&= \int_{\Gamma g \in P(M)} \left(\int_{\Delta^n} (p \operatorname{str} \tilde{f} g \sigma_0)^* \Omega_N \right) dh_M \\
&= \int_{\Gamma g \in P(M)} \left(\int_{\Delta^n} (\operatorname{str} \tilde{f} g \sigma_0)^* \Omega_{\mathbb{H}^n} \right) dh_M \\
&< m_A(\operatorname{Vol}(\sigma_0) - 2\varepsilon + \delta) + (\operatorname{Vol} M - m_A)(\operatorname{Vol} \sigma_0 + \delta) \\
&= \operatorname{Vol}(M)(\operatorname{Vol}(\sigma_0) + \delta) - 2m_A\varepsilon.
\end{aligned}$$

Letting $\delta < (\varepsilon m_A)/\operatorname{Vol} M$, we obtain

$$\langle \operatorname{str} \tilde{f}_*(\operatorname{smear}(\sigma_0)), \Omega_N \rangle < \operatorname{Vol} M \operatorname{Vol} \sigma_0 - \varepsilon m_A. \tag{3.1}$$

The map $f : M \to N$ is a homotopy equivalence, so that $f_*([M]) = [N]$. By Theorem 3.4, M and N must have the same volume. Since $\zeta(\sigma_0)$ represents $|\operatorname{Vol}(\sigma_0)|[M]$, it follows that $\operatorname{str}(f_*(\zeta(\sigma_0)))$ represents $|\operatorname{Vol}(\sigma_0)|[N]$.

On other hand, since $\operatorname{Vol}(M) = \operatorname{Vol}(N)$, the equation (3.1) implies that $\operatorname{str}(f_*(\zeta(\sigma_0)))$ represents $\lambda[N]$, with $\lambda < |\operatorname{Vol} \sigma_0| - \varepsilon m_A/\operatorname{Vol} M$. This is a contradiction.

□

We now know how the boundary map acts on simplices of maximal volume. But we need more precise geometric information. This is going to be provided by Theorem 3.12, proved by Haagerup and Munkholm [10]. Before stating it, regular simplices and some of their properties have to be introduced.

Definition 3.9. A simplex σ in $\mathbb{H}^n$ with vertices $v_0, \ldots, v_n$ is ***regular*** if for each $i < j$, there exists an isometry of $\mathbb{H}^n$ fixing $v_0, \ldots, \hat{v}_i, \ldots, \hat{v}_j, \ldots, v_n$ and swapping v_i and v_j.

Lemma 3.10. *In the upper half-space model, let $v_0, \ldots, v_n$ be an ideal simplex with $v_0 = \infty$. Then the simplex is regular if and only if $v_1, \ldots, v_n$ span a regular Euclidean simplex.*

Proof. Straightforward from the definition of a regular simplex. □

Lemma 3.11. $\mathrm{Isom}(\mathbb{H}^n)$ *acts transitively on the set of regular ideal simplices.*

Proof. This follows from Proposition 1.1. □

Theorem 3.12. *An ideal simplex in $\mathbb{H}^n$ has maximal volume if and only if it is regular.*

Gromov's proof of Mostow rigidity. To simplify the notation, we carry out the proof for $n = 3$. It is exactly the same proof in higher dimensions. Let $v_0, \ldots, v_3$ be vertices in S^2 spanning an ideal simplex of maximal volume in $\mathbb{H}^3$. Then $h(v_0), \ldots, h(v_3)$ span an ideal simplex of maximal volume which must be regular by Theorem 3.12. By Lemma 3.11, up to an isometry, we can assume that h fixes $v_0, \ldots, v_3$. Now work in the upper half-space model and suppose that $v_0 = \infty$. We have to show that h is the identity.

We first claim that h fixes a tiling of the horizontal plane by equilateral triangles isometric to $v_1, \ldots, v_3$. To see this, use Lemma 3.10 to observe that $v_1, \ldots, v_3$ is an equilateral triangle. Then let v_1' be the reflection of v_1 through the subsimplex v_0, v_2, v_3, as shown in Figure 3.2. Since h is injective, fixes v_0, v_2, v_3, and sends regular simplices to regular simplices, it must also fix v_1' (here we again use Lemma 3.10). Argue similarly for the other vertices of this tiling.

Second, h fixes a finer tiling of the horizontal plane by equilateral triangles. Use the reflections v_0' and v_0'' of v_0 shown in Figure 3.2. By the same argument as above, h must fix v_0' and v_0''. Then observe that v_2, v_0', v_0'' form a new equilateral triangle smaller than the former. Inducting these two steps implies that h fixes a dense tiling of S^2. By continuity, h is the identity.

□

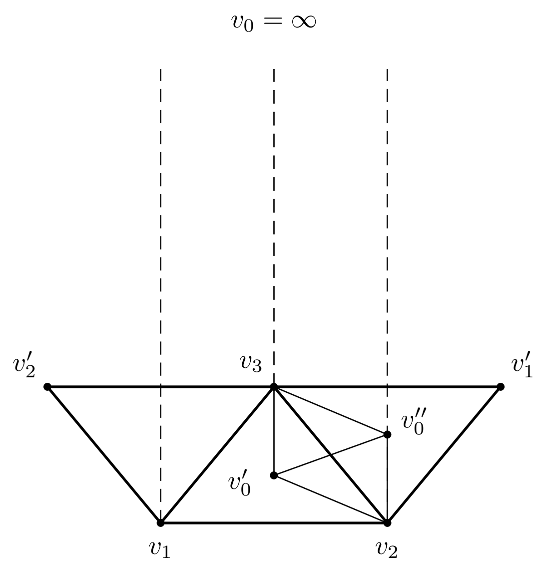


Figure 3.2: Tiling of $\partial\mathbb{H}^n$ by iterated reflections

Chapter 4

Minimal entropy

This chapter contains a third proof of Mostow rigidity that is again very different from the preceding proofs, except for the use of the boundary map. The main result characterizes locally symmetric metrics amongs metrics on negatively curved manifolds that are related by a homotopy equivalence. This characterization is given in terms of an invariant known as the *entropy* of the manifold. Mostow's rigidity theorem will then be a straightforward corollary.

Although we give a complete proof only for real hyperbolic spaces, the employed methods extend rather easily to complex and quaternionic hyperbolic spaces. Moreover, the theorem we prove is a special case of a stronger theorem from which many corollaries can be deduced (see [4] and [3]). Thus, the methods employed are very fruitful. Nevertheless, we will restrict our attention to Mostow rigidity.

While the techniques used in the preceding chapters were mainly analytical and homological, this chapter relies on differential geometry, construction of measures and multi-variable calculus. In addition, this proof provides an explicit construction of the desired isometry.

Throughout this section, (M, g) will be a compact connected negatively curved Riemannian n-manifold. The main references for this chapter are [4], [3] and [8].

4.1 Entropy and outline

We first introduce the entropy of a metric, then state the theorem and outline the proof. Details are filled in in subsequent sections.

Intuitively speaking, the entropy measures how quickly the volume of a ball grows.

Definition 4.1. Let (X, g) be a Riemannian manifold. Let $B_p(R)$ be a ball of radius R about p in the universal cover $\tilde{X}$ of X. The ***(volumic) entropy*** of (X, g) is

$$h(g) = \lim_{R \to \infty} \frac{1}{R} \log(\mathrm{Vol}(B_p(R))),$$

for some $p \in \tilde{X}$.

Manning [14] proved that this definition is meaningful in the following sense.

Proposition 4.2. *The entropy of a compact Riemannian manifold* (X, g) *always exists and is independent of the choice of basepoint.*

Sketch of proof. We sketch the independence from the basepoint. Since X is compact, it has a fundamental domain in $\tilde{M}$ of finite diameter D. Let p_0, p_1 be two basepoints in $\tilde{X}$. Moving either point by an isometry, we can assume that p_0 and p_1 lie in the same copy of the fundamental domain and so are at distance at most D from each other. It is now easy to see that that $\mathrm{Vol}(B_{p_0}(r - D) \leq \mathrm{Vol}(B_{p_1}(r)) \leq \mathrm{Vol}(B_{p_0}(r + D))$, so that the limits are independent of the choice of basepoint, provided they exist. □

The flaw in this invariant is that it is not invariant under rescaling and thus distinguishes metrics that are essentially the same. Precisely, for any $\lambda > 0$, we have $h(\lambda g) = \frac{1}{\lambda} h(g)$. To see this, notice that $\mathrm{Vol}(B_p(R), \lambda g) = \mathrm{Vol}(B_p(R/\lambda), g)$ and so

$$\begin{aligned} h(\lambda g) &= \lim_{R \to \infty} \frac{1}{R} \log(\mathrm{Vol}(B_p(R), \lambda g)) = \frac{1}{\lambda} \lim_{R \to \infty} \frac{\lambda}{R} \log(\mathrm{Vol}(B_p(R/\lambda), g)) \\ &= \frac{1}{\lambda} h(g). \end{aligned}$$

Therefore, instead of just considering the entropy, we will consider the functional $\mathrm{Vol}(M, g) h^n(g)$. It is now straightforward to see that this is invariant under rescaling of the metric.

Example 4.3. Entropy of a hyperbolic manifold. The computation is simplest when using the hyperboloid model and assuming that $p = (0, \ldots, 0, 1)$. Use the usual parametrization

$$(x_1, \ldots, x_n) \longmapsto (x_1, \ldots, x_n, \sqrt{1 + x_1^2 + \ldots + x_n^2})$$

and hyperspherical coordinates to find

$$\begin{aligned} \mathrm{Vol}(B_p(R)) &= \int_{\|x\| \le \sinh R} \frac{dx_1 \ldots dx_n}{\sqrt{1 + x_1^2 + \ldots + x_n^2}} \\ &= c_n \int_0^{\sinh R} \frac{r^{n-1}}{\sqrt{1+r^2}}\, dr \\ &= c_n \int_0^R \sinh^{n-1}(u)\, du \\ &\sim e^{(n-1)R}. \end{aligned}$$

Therefore, the entropy of a hyperbolic manifold is $n - 1$.

The manifolds of interest in this theorem are locally symmetric spaces.

Definition 4.4. A ***locally symmetric space*** is a homogeneous connected Riemannian manifold X such that for any $p \in X$, there is a symmetric neighborhood U of p on which the geodesic symmetry map is a local isometry. The geodesic symmetry map at p fixes the point p and reverses all geodesics through that point.

The theorem this chapter focuses on is the following, due to Besson, Courtois and Gallot (see [4]).

Theorem 4.5. *Let (M, g), (N, g_0) be compact, connected, negatively curved Riemannian n-manifolds with $n \ge 3$. Suppose that g_0 is a locally symmetric metric and that there exists a homotopy equivalence $f : M \to N$. Then*

1. $h^n(g)\,\mathrm{Vol}(M, g) \ge h^n(g_0)\,\mathrm{Vol}(N, g_0)$,
2. $h(g) = h(g_0)$ *and* $\mathrm{Vol}(M, g) = \mathrm{Vol}(N, g_0)$ *if and only if (M, g) is isometric to (N, g).*

The second statement can be rephrased as follows. There is equality if and only if the manifolds are isometric *up to rescaling of the metrics*.

Mostow's rigidity theorem follows as an easy corollary.

Proof of Mostow rigidity. Since both manifolds M and N are assumed to be hyperbolic, they are locally symmetric, thus the inequality of 4.5.1 holds in both ways and so equality holds. Theorem 4.5.2 then implies that M and N are isometric up to rescaling of the metrics. □

We now outline the proof of Theorem 4.5.

1. The boundary map

Exactly as in Chapters 2 and 3, we use the $\pi_1(M)$-equivariant lift $\tilde{f} : \tilde{M} \to \tilde{N}$ and the boundary map $h = \partial\tilde{f} : \partial\tilde{M} \to \partial\tilde{N}$. Nevertheless, unlike in Chapter 2, we need only the fact that h is a homeomorphism. No heavy analysis facts will be quoted. Instead, we use tools such as Patterson-Sullivan measures, the barycenter construction and the map h to define a map from $\tilde{M}$ to $\tilde{N}$ which turns out to be a $\pi_1(M)$-equivariant isometry.

2. Patterson-Sullivan measures

Denote $\mathcal{M}(\partial\tilde{M})$ denote the space of Borel measures on $\partial\tilde{M}$. We construct a map

$$\begin{array}{ccc} \tilde{M} & \longrightarrow & \mathcal{M}(\partial\tilde{M}) \\ y & \longmapsto & \mu_y \end{array}$$

that is $\pi_1(M)$-equivariant. The measure μ_y will arise from a limiting process involving measures of the form

$$\nu_{y,s} = \frac{\sum_{\gamma\in\Gamma} e^{-sd(y,\gamma.p_0)}\delta_{\gamma.p_0}}{\sum_{\gamma\in\Gamma} e^{-sd(p_0,\gamma.p_0)}},$$

for some basepoint $p_0 \in M$. The series converges for $s > h(g)$ and the denominator can be assumed (for mathematical reasons) to diverge at $s = h(g)$. From this fact and a compactness argument, some subsequence ν_{y,s_i} converges to a measure concentrated on $\partial\tilde{M}$. The fact that Γ is a lattice ensures that the resulting measure is nonatomic.

The Radon-Nikodym derivative of these measures has the following nice form. For y, $y' \in \tilde{M}$,

$$\frac{d\mu_y}{d\mu_{y'}}(\theta) = e^{-h(g)B_{y'}(y,\theta)}, \tag{4.1}$$

where $B_{y'}(\cdot, \theta)$ is the Busemann function of θ centered at y' (see Section 1.1 for the basics of Busemann functions).

3. The barycenter method

Let X be the universal cover of a complete nonpositively curved manifold and let Γ be a discrete group acting cocompactly by isometries on X. Suppose that ∂X is endowed with a nonatomic measure λ. Just as Busemann functions formalize the concept of distance to infinity, the barycenter of λ captures the idea of the "closest" point to the boundary at infinity with respect to this measure.

Fix an origin $o \in X$. The barycenter of λ, written $\operatorname{bar} \lambda$, is the unique minimum x of the functional

$$\int_{\tilde{X}} B_o(x, \theta)\, d\lambda(\theta).$$

This defines a Γ-equivariant map

$$\begin{array}{rcl} \mathcal{M}(\partial X) & \longrightarrow & X \\ \lambda & \longmapsto & \operatorname{bar} \lambda. \end{array}$$

4. The natural map and its properties

Given steps 2 and 3, the definition of the natural map is now quite natural. Given $y \in \tilde{M}$, construct the Patterson-Sullivan measure μ_y on $\partial \tilde{M}$, push it forward to $\partial \tilde{N}$ by h and take the barycenter.

$$\begin{array}{rclclcl} \tilde{F} : \tilde{M} & \to & \mathcal{M}(\partial \tilde{M}) & \to & \mathcal{M}(\partial \tilde{N}) & \to & \tilde{N} \\ y & \mapsto & \mu_y & \mapsto & h_* \mu_y & \to & \operatorname{bar}(h_* \mu_y) = \tilde{F}(y). \end{array}$$

As a consequence of the preceding steps, $\tilde{F}$ is $\pi_1(M)$-equivariant and so descends to a map F. Since all the following estimates are pointwise, we will not distinguish F and $\tilde{F}$ any more. Since F induces the same isomorphism of fundamental groups as f, the two maps are homotopic. Theorem 4.5 will directly follow from the following proposition.

Proposition 4.6. *The function F is at least C^1 and satisfies*

1. $|\operatorname{Jac} F(y)| \leq \frac{h^n(g)}{h^n(g_0)}$ *for all $y \in M$,*

2. *equality holds if and only if F is a homothety of ratio $h(g)/h(g_0)$.*

Proof of Theorem 4.5. Let ω (resp. ω_0) be the volume form on (M, g) (resp. (N, g_0)). Since F is homotopic to f, it has degree one, so Proposition 4.6.1 implies that

$$\begin{aligned}\mathrm{Vol}(N, g_0) &= \int_Y \omega_0 = \int_X F^*\omega_0 = \int_X |\operatorname{Jac} F|\omega \\ &\leq \frac{h^n(g)}{h^n(g_0)} \int_X \omega = \frac{h^n(g)}{h^n(g_0)} \mathrm{Vol}(M, g).\end{aligned}$$

This implies the first statement. Now, if equality holds, then $|\operatorname{Jac} F(y)| = h^n(g)/h^n(g_0)$ so by Proposition 4.6.2, F is a homothety. Rescaling the metrics so that $h(g) = h(g_0)$ implies that $\operatorname{Jac} F = I_n$. In other words, F is an isometry. □

4.2 Patterson-Sullivan measures

Let X be the universal cover of a complete nonpositively curved manifold and let Γ be a discrete group acting cocompactly by isometries on X. We fix once and for all a basepoint p_0 in X. The purpose of this section is to construct a family $(\mu_y)_{y \in X}$ of probability measures such that the map

$$\begin{aligned} X &\longrightarrow \mathcal{M}(X) \\ y &\longmapsto \mu_y \end{aligned}$$

is Γ-equivariant and satisfies

$$\frac{d\mu_y}{d\mu_{y'}}(\theta) = e^{-h(g)B_{y'}(y,\theta)} \quad \text{for all } \theta \in \partial X.$$

Intuitively, the Patterson-Sullivan measure at $y \in X$ measures the visual density at infinity of the orbit $\Gamma.p_0$, as seen from the point y. This will be obtained by a limiting process. In hyperbolic spaces, the visual size of objects decreases exponentially as the object moves further away. Thus, for each $\gamma \in \Gamma$, we place a Dirac measure at $\gamma.p_0$ and we scale it by a factor $e^{-sd(y,\gamma.p_0)}$ for some adequately chosen $s > 0$. For convergence purposes, we divide the resulting sum by another infinite sum so as to define

$$\nu_{y,s} = \frac{\sum_{\gamma\in\Gamma} e^{-sd(y,\gamma.p_0)}\delta_{\gamma.p_0}}{\sum_{\gamma\in\Gamma} e^{-sd(p_0,\gamma.p_0)}}.$$

These series will shortly be seen to converge when $s > h(g)$. For x, $y \in X$, define

$$g_s(y,z) = \sum_{\gamma\in\Gamma} e^{-sd(y,\gamma.z)}.$$

Let $a_n = \#\{\Gamma.p_0 \cap (B_{p_0}(n) - B_{p_0}(n-1))\}$ and observe that

$$g_s(p_0,p_0) \;\sim\; \sum_{n=1}^{\infty} a_n e^{-sn}$$

Let $b_n = \mathrm{Vol}(B_{p_0}(n))$. The cocompactness of Γ implies that $a_n \sim b_n - b_{n-1}$. Since $s > h(g)$, there is an $\varepsilon > 0$ such that $s - \frac{1}{n}\log b_n > \varepsilon$ for large n. Therefore

$$\sum_{n=1}^{\infty} b_n e^{-sn} = \sum_{n=1}^{\infty} e^{-(s-\frac{1}{n}\log b_n)n}$$

and this last expression converges. This implies the convergence of the series $\sum_{n=1}^{\infty} a_n e^{-sn}$. The same argument shows the divergence of this series when $s < h(g)$. Using the triangle inequality, it is easily seen that $g(p_0,p_0)$ converges if and only if $g(y,z)$ converges for all y, $z \in X$.

We do not know whether the series $g_s(y,z)$ converges or diverges when $s = h(g)$. In fact, the series can be modified so that it diverges at $s = h(g)$ and still converges when $s < h(g)$. We refer the reader to [8] or [18] where this is explained thoroughly. The idea is to multiply each term in the series $g(y,p_0)$ by $u(e^{d(y,\zeta)})$, where $\zeta \in \bar{X}$ and $u : [0,\infty) \to [0,\infty)$ is an increasing function that grows just slightly faster than the identity function. For simplicity of our discussion, we simply assume that the series $g_s(\cdot,\cdot)$ diverges when $s = h(g)$.

Define

$$\nu_{s,y} = \frac{\sum_{\gamma\in\Gamma} e^{-sd(y,\gamma.p_0)}\delta_{\gamma.p_0}}{\sum_{\gamma\in\Gamma} e^{-sd(p_0,\gamma.p_0)}}.$$

The family of Γ-equivariant maps $\{y \mapsto \nu_{s,y} : y \in X, s \in (h(g), h(g)+1]\}$ is a subset of $C(X, \mathcal{M}(\bar{X}))$. It is an exercise to show that this family is equicontinuous and uniformly bounded on compact sets, so by the Arzela-Ascoli theorem, this family is a relatively compact subset of $C(X, \mathcal{M}(\bar{X}))$ endowed with the topology of uniform convergence on compact sets. It follows that for all $y \in X$, there is a sequence $\nu_{s_i,y}$ converging to a measure ν_y. The divergence of the series $g_{h(g)}(p_0,p_0)$ implies that ν_y lies in $\mathcal{M}(\partial X)$. The measure ν_y is concentrated on the set of accumulation points of $\Gamma.p_0$. Since Γ is cocompact, the measure ν_y is supported on the whole boundary ∂X.

To see that the measure ν_y is nonatomic, suppose for the sake of contradiction that there exists some $\xi \in \partial X$ with $\mu_y(\{\xi\}) > 0$. Since M is cocompact, $\pi_1(M)$ must contain a hyperbolic element γ such that its axis does not have ξ as endpoint. Equivariance of ν_y and the fact that the set $\{\gamma^m.\xi : m \in \mathbb{Z}\}$ is infinite imply that ν_y has infinite mass, which is absurd.

We finally prove the property of change of measure. Namely, for y, $y' \in X$ and $\theta \in \partial X$, we show that

$$\frac{d\nu_y}{d\nu_{y'}}(\theta) = e^{-sB_{y'}(y,\theta)}. \tag{4.2}$$

Consider a point $\gamma.p_0$ close to θ. When computing the ratio $\nu_{y,s}/\nu_{y',s}$, the coefficient in front of each Dirac measure $\delta_{\gamma.p_0}$ close to θ is

$$\frac{e^{-sd(y,\gamma.p_0)}}{e^{-sd(y',\gamma.p_0)}} = e^{-s(d(y,\gamma.p_0)-d(y',\gamma.p_0))} \simeq e^{-sB_{y'}(y,\theta)}.$$

This is illustrated on Figure 4.1. As $s_i \to h(g)$, only terms near the boundary count. This proves (4.2).

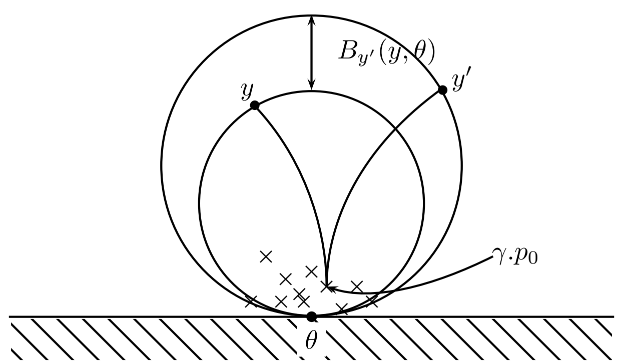


Figure 4.1: Change of measure property

Renormalizing the measures ν_y, we obtain probability measures on ∂X via the map

$$\begin{array}{rcl} X & \longrightarrow & \mathcal{M}(\partial X) \\ y & \longmapsto & \mu_y = \dfrac{1}{\nu_y(\partial X)} \cdot \nu_y. \end{array}$$

The property of change of measures only changes up to a multiplicative constant. Thus

$$\frac{d\mu_y}{d\mu_{y'}}(\theta) = c_{y,y'} e^{-sB_{y'}(y,\theta)}, \tag{4.3}$$

where $c_{y,y'} = \nu_{y'}(\partial X)/\nu_y(\partial X)$.

4.3 The barycenter

Let X again be the universal cover of a negatively curved manifold. As said above in the outline, the barycenter of a nonatomic measure on $\partial\tilde{X}$ is in some sense the closest point in X to the boundary with respect to that measure. This part uses Riemannian geometry and relies on the understanding of the curvature of horospheres. The main result of this section asserts the existence and uniqueness of the barycenter.

Proposition 4.7. *Let $\lambda \in \mathcal{M}(\partial X)$ be a nonatomic measure. Then the function*

$$r_\lambda(x) = \int_{\partial X} B_o(x,\theta)\, d\lambda(\theta)$$

has a unique minimum in X, denoted by $\mathrm{bar}(\lambda)$.

First observe that the map

$$\begin{array}{rcl} \mathcal{M}(\partial X) & \longrightarrow & X \\ \lambda & \longmapsto & \mathrm{bar}\,\lambda \end{array}$$

is Γ-equivariant, that is, $\mathrm{bar}(\gamma_*\lambda) = \gamma.\mathrm{bar}(\lambda)$. Write $x_0 = \mathrm{bar}(\lambda)$. One has to show that $\gamma.x_0$ minimizes $r_{\gamma_*\lambda}$. To see this, recall that $B_o(\cdot,\theta)$ and $B_{o'}(\cdot,\theta)$ differ by a constant $c = c(\theta)$. Then

$$\begin{aligned} r_{\gamma_*\lambda}(\gamma.x) &= \int_{\partial X} B_o(\gamma.x,\theta)\, d(\gamma_*\lambda)(\theta) = \int_{\partial X} B_o(\gamma.x,\gamma.\theta)\, d\lambda(\theta) \\ &= \int_{\partial X} B_{\gamma^{-1}o}(x,\theta)\, d\lambda(\theta) \\ &= \int_{\partial X} B_o(x,\theta)\, d\lambda(\theta) + \int_{\partial X} c(\theta)\, d\lambda(\theta). \end{aligned}$$

and notice that this expression is minimal when $x = x_0$.

Existence and uniqueness of the barycenter follow from the strict convexity of the function $x \mapsto \int_{\partial X} B_o(x,\theta)\, d\lambda(\theta)$. This will come from the fact that the Hessian of Busemann functions is positive definite for almost all vectors, which is a consequence of the next two propositions.

Proposition 4.8. *For any $\theta \in S^{n-1} \simeq \partial \mathbb{H}^n$, the Hessian of the Busemann function $B(\cdot,\theta)$ at p equals the second fundamental form of the horosphere at θ through p.*

Proof. Fix $p \in \mathbb{H}^n$, $\theta \in S^{n-1}$ and for simplicity write $B(\cdot) = B_o(\cdot,\theta)$. Recall that the second fundamental form of a hypersurface S is the symmetric form given by

$$II(u,u) = -g(\nabla_u \nu, u), \qquad \text{for all } u \in TS,$$

where ν is a vector field normal to S. The Hessian of B at p is $H_p(B)(v,v) = (\nabla_v dB)(v)$, where $v \in T_p\mathbb{H}^n$.

Write $S = HS(p,\theta)$. Pick a chart (U,φ) around p such that $g_{ij}(p) = \delta_{ij}$. Let ν be the vector field normal to the horosphere $HS(p,\theta)$. Up to an isometry of $\mathbb{R}^n$, we can assume that (U,φ) is such that $\nu(p) = -\frac{\partial}{\partial x^n}$.

For $q \in HS(p,\theta)$, write $\nu(q) = a_k(q)\frac{\partial}{\partial x^k}$ and notice that at point p,

$$a_k(p) = \begin{cases} 0, & 1 \le k \le n-1 \\ -1, & k = n \end{cases}.$$

Also notice that $dB = a_k dx^k$.

We are going to prove that $II(u,u) = H_p(B)(u,u)$ for $u = \frac{\partial}{\partial x^l} \in T_pS$, with $l = 1,\ldots,n-1$.

By definition, $\nabla_u dh = C(u \otimes \nabla dh)$, where C is the contraction operator.

$$\begin{aligned} \nabla dB &= \nabla(a_k dx^k) = da^k \otimes dx^k + a_k \nabla dx^k \\ &= da^k \otimes dx^k - a_k \Gamma^k_{ij} dx^i \otimes dx^j. \end{aligned}$$

Since $u = \frac{\partial}{\partial x^l}$, one has $dx^i(u) = \delta_{il}$ and so

$$\begin{aligned} (\nabla_u dB)(u) &= da^k(u) dx^k(u) - a_k \Gamma^k_{ij} dx^i(u) dx^j(u) \\ &= da^l(u) + \Gamma^n_{ll}(p). \end{aligned}$$

On the other hand,

$$\nabla \nu = \nabla(a_k \frac{\partial}{\partial x^k}) = da^k \otimes \frac{\partial}{\partial x^k} + a_k \Gamma^i_{kj} \frac{\partial}{\partial x^i} \otimes dx^j$$

so that

$$\begin{aligned}\nabla_u \nu = C(u \otimes \nabla \nu) &= da^k(u)\frac{\partial}{\partial x^k} + a_k \Gamma^i_{kj}\frac{\partial}{\partial x^i} dx^j(u) \\ &= da_k(u)\frac{\partial}{\partial x^k} - \Gamma^i_{nl}(p)\frac{\partial}{\partial x^i},\end{aligned}$$

Thus

$$\begin{aligned} g(\nabla_u \nu, u) &= g(da_k(u)\frac{\partial}{\partial x^k} - \Gamma^k_{nl}(p)\frac{\partial}{\partial x^k}, \frac{\partial}{\partial x^l}) \\ &= da^l(u) - \Gamma^l_{nl}(p) = da^l(u) + \Gamma^n_{ll}(p) \\ &= H_p(B)(u,u). \end{aligned}$$

□

Proposition 4.9. *Let M be a compact Riemannian manifold of constant sectional curvature $\kappa = -1$ and X its universal cover. Then horospheres in X have positive sectional curvatures equal to 1.*

Proof. Fix $\theta \in \partial X$ and write $B = B(\cdot, \theta)$. Let ν be the gradient vector field of B and let $\gamma_p(t)$ be the geodesic curve with initial tangent vector $\nu(p)$. Let $w \in \nu(p)^\perp$ and define a vector field along γ_p

$$Y(t) = d(\gamma_p)_t w.$$

This is a Jacobi field. For any $q \in HS(q, \theta)$, the geodesics γ_p and γ_q converge to the same point in ∂X when t tends to $-\infty$. Therefore

$$\lim_{t \to -\infty} \|Y(t)\| = 0. \tag{4.4}$$

Let $w(t)$ be the parallel translation of w along γ_p. Choose an orthonormal frame at p containing the vectors $\nu(p)$, w and parallel translate it along γ_p. Use this with (4.4) to solve the Jacobi equation. This yields $Y(t) = e^t w(t)$. Also observe that $\frac{\nabla Y}{dt}\Big|_{t=0} = \nabla_\nu Y = \nabla_w \nu$. Therefore

$$II(w,w) = \langle \nabla_w \nu, w \rangle = \frac{1}{2}\frac{d}{dt}\Big|_{t=0} \langle Y(t), Y(t) \rangle = g_0(w,w) = 1.$$

□

Corollary 4.10. *Under the same hypotheses, for all $p \in X$ and $u, v \in T_pX$,*

$$H_p(B)(u,v) = g_0(u,v) - dB(u)dB(v)$$

Proof. The preceding proof and Proposition 4.8 shows that $H_p(B)(v,v) = g_0(v,v)$ for all $v \in TS$. An easy computation implies that $H_p(B)(u,u) = -dB(u)^2$. Using that $dB(v) = 0$ for all $v \in TS$, the result easily follows. ☐

Proof of Proposition 4.7. To prove that the function r_λ has an absolute minimum, it suffices to check that it is strictly convex and that $r_\lambda(x)$ tends to infinity as x approaches the boundary ∂X.

It follows from Corollary 4.10 that $(\nabla_u B_{(x,\theta)})(u) = 0$ if and only if θ is an endpoint of the geodesic line with tangent vector u. Moreover, Proposition 4.9 implies that $\nabla_u B_{(x,\theta)}(u)$ is positive definite. Thus, since λ is nonatomic, the bilinear form

$$(\nabla_u r_\lambda)(u) = \int_{\partial X} (\nabla_u B_{o_{(x,\theta)}})(u)\, d\lambda(\theta)$$

is positive definite and so r_λ is strictly convex.

It remains to check that $r(x)$ goes to infinity when x tends to $\theta \in \partial X$. This uses the convexity of Busemann functions and standard measure-theoretic arguments. See [3, Appendix A] or [8] for more details. ☐

4.4 Computation of the Jacobian

Recall that the natural map $F(y) = \mathrm{bar}(h_* d\mu_y)$ is implicitly defined by the vector-valued equation

$$\int_{\partial \tilde{N}} dB_{o_{(F(y),\theta)}}(\cdot)\, d(h_*\mu_y)(\theta) = 0. \tag{4.5}$$

Using (4.1), this is equivalent to

$$\int_{\partial \tilde{M}} dB_{o_{(F(y),h(\alpha))}}(\cdot) e^{-h(g)B(\alpha,y)}\, d\mu_{p_0}(\alpha) = 0. \tag{4.6}$$

It is natural to use the implicit function theorem to show the regularity of F and the properties of Proposition 4.6. Let $(e_i(z))_{i=1,\dots,n}$ be a family of frames on $T_z\tilde{N}$ depending smoothly on z. Define

$$\begin{aligned} G : \tilde{M} \times \tilde{N} &\longrightarrow \mathbb{R}^n \\ (y,z) &\longmapsto (G_1(y,z), \dots, G_n(y,z)), \end{aligned}$$

where
$$G_i(z) = \int_{\partial\tilde{M}} dB_{o_{(z,h(\alpha))}}(e_i(z))e^{-h(g)B(\alpha,y)}\, d\mu_{p_0}(\alpha).$$
To ckeck the nondegeneracy condition, carefully differentiate G with respect to z. For $v \in T_y\tilde{M}$ and $G(y,z) = 0$, one finds

$$(D_zG_i)_{(y,z)}(v) = \int_{\partial\tilde{M}} \nabla_v dB_{o_{(z,h(\alpha))}}(e_i(z))\, d\mu_y(\alpha)$$

By the same argument as in the proof of Proposition 4.7, one concludes that D_zG is invertible. Therefore, by the implicit function theorem, the natural map F is C^1. To estimate the Jacobian of F, differentiate the equation $G(y,F(y)) = 0$. Then for all $u \in T_y\tilde{M}$ and $v \in T_{F(y)}\tilde{N}$,

$$\int_{\partial\tilde{M}} \nabla_{dF_y(u)} dB_{o_{(F(y),h(\alpha))}}(v)\, d\mu_y(\alpha) = h(g) \int_{\partial\tilde{M}} dB_{o_{(F(y),h(\alpha))}}(v) dB_{(\alpha,y)}(u)\, d\mu_y(\alpha). \tag{4.7}$$

Using the Cauchy-Schwarz inequality, we obtain

$$\begin{aligned} &\left| \int_{\partial\tilde{M}} \nabla_{dF_y(u)} dB_{o_{(F(y),h(\alpha))}}(v)\, d\mu_y(\alpha) \right| \\ &\leq h(g) \left(\int_{\partial\tilde{M}} dB^2_{o_{(F(y),h(\alpha))}}(v)\, d\mu_y(\alpha) \right)^{1/2} \left(\int_{\partial\tilde{M}} dB^2_{(\alpha,y)}(u)\, d\mu_y(\alpha) \right)^{1/2}. \end{aligned} \tag{4.8}$$

This prompts us to define symmetric endomorphisms $K_{F(y)}$, $H_{F(y)}$ on $T_{F(y)}\tilde{N}$ via the bilinear forms

$$\begin{aligned} g_0(K_{F(y)} \circ dF_y(u), v) &= \int_{\partial\tilde{M}} \nabla_{dF_y(u)} dB_{o_{(F(y),h(\alpha))}}(v)\, d\mu_y(\alpha) \\ g_0(H_{F(y)}(v), v) &= \int_{\partial\tilde{M}} dB^2_{o_{(F(y),h(\alpha))}}(v)\, d\mu_y(\alpha). \end{aligned}$$

Thus (4.8) can be rewritten as

$$|g_0(K \circ dF_y(u), v)| \leq h(g) \cdot g_0(H(v), v) \cdot \left(\int_{\partial\tilde{M}} dB^2_{(\alpha,y)}(u)\, d\mu_y(\alpha) \right)^{1/2}. \tag{4.9}$$

From now on, omit subscripts for simplicitly and write $\langle \cdot, \cdot \rangle_0$ instead of $g_0(\cdot,\cdot)$. Since Proposition 4.6 is trivial when dF_y is not invertible, we can assume that it is invertible. Notice that K is invertible, since the bilinear form $g_0(K\cdot,\cdot)$ equals the Hessian of B_0 which is positive definite. This ensures that the map $K \circ dF_y$ is invertible. In the following, linear maps will be considered as matrices and a good choice of basis will make the computation easier.

Let (u_i) be a basis of $T_{F(y)}\tilde{N}$ diagonalizing H and let (v_i) be the orthonormal basis of $T_y\tilde{M}$ obtained by applying the Gram-Schmidt process to the basis $((K \circ dF_y)^{-1}(u_i))$. Then the matrix of $K \circ dF_y$ is upper triangular. Use (4.8) and the inequality of geometric and arithmetic means to find

$$\begin{aligned} \det(K \circ dF_y) &= \prod_{i=1}^{n} \langle K \circ dF_y(v_i), u_i\rangle_0 \\ &\leq h^n(g)(\det H)^{1/2} \left(\frac{1}{n} \int_{\partial\tilde{M}} \sum_{i=1}^{n} dB^2_{(y,\alpha)}(v_i)\, d\mu_y(\alpha) \right)^{n/2} \\ &= h^n(g)(\det H)^{1/2} n^{-n/2}, \end{aligned} \tag{4.10}$$

where the last equality comes from the fact that $\sum_{i=1}^{n} dB^2_{(y,\alpha)}(v_i) = \|dB^2_{(y,\alpha)}\|^2 = 1$ and that μ_y is a probability measure.

Now we restrict our attention to real hyperbolic spaces. The general proof for complex and quaternionic hyperbolic spaces is given in [4] and [8]. Integrating the formula of Corollary 4.10 yields

$$H = I - K.$$

Thus (4.10) becomes

$$|\mathrm{Jac}\, F(y)| \leq \frac{h^n(g)}{n^{n/2}} \cdot \frac{(\det H)^{1/2}}{\det(I - H)}. \tag{4.11}$$

Also observe that

$$\begin{aligned} \mathrm{Tr}(H) &= \sum_{i=1}^{n} \langle Hu_i, u_i\rangle_0 = \sum dB^2_{o_{(F(y),h(\alpha))}}(u_i)\, d\mu_y(\alpha) \\ &= 1, \end{aligned}$$

for the same reason as in (4.10).

The next linear algebra result is the cornerstone of the proof. Notice that it is the only step requiring that $n \geq 3$.

Lemma 4.11. *Let H be a symmetric definite positive matrix with trace 1. Assume that $n \geq 3$. Then*

$$\frac{\det(H)^{1/2}}{\det(I-H)} \leq \left(\frac{\sqrt{n}}{n-1}\right)^n,$$

with equality if and only if $H = I/n$. Moreover, this is false when $n = 2$.

Proof. The proof uses Lagrange multipliers. See [3, Appendix B]. □

This result and (4.11), together with the fact that $h(g_0) = n-1$, imply that

$$|\operatorname{Jac} F(y)| \leq \frac{h^n(g)}{h^n(g_0)} \tag{4.12}$$

There remains the equality case. The second part of Lemma 4.11 implies that $H = \frac{1}{n}I$ and $K = \frac{n-1}{n}I$. Thus (4.8) becomes

$$|\langle dF_y(u), v\rangle_0| \leq \frac{h(g)}{h(g_0)} n^{1/2} \|v\|_0 \left(\int_{\partial\tilde{M}} dB_{(\alpha,y)}(u)\, d\mu_y(\alpha)\right)^{1/2}. \tag{4.13}$$

Taking the supremum over all v with $\|v\| = 1$ yields

$$\|dF_y(u)\|_0 \leq \frac{h(g)}{h(g_0)} n^{1/2} \left(\int_{\partial\tilde{M}} dB_{(\alpha,y)}(u)\, d\mu_y(\alpha)\right)^{1/2}. \tag{4.14}$$

Let $L = dF_y^* \circ dF_y$. It is a normal endomorphism of $T_y\tilde{M}$. Let (w_i) an orthonormal basis diagonalizing L. We are going to show that $\det L = (\frac{\operatorname{Tr} L}{n})^n$. The equality case of the inequality of arithmetic and geometric means will then imply that $L = \frac{h^2(g)}{h^2(g_0)} I$. It follows from (4.14) that

$$\operatorname{Tr} L = \sum_{i=1}^n \langle dF_y(w_i), dF_y(w_i)\rangle \leq \frac{h^2(g)}{h^2(g_0)} \cdot n. \tag{4.15}$$

Thus

$$\frac{h^{2n}(g)}{h^{2n}(g_0)} = \det L \leq \left(\frac{\operatorname{Tr} L}{n}\right)^n \leq \frac{h^{2n}(g)}{h^{2n}(g_0)}.$$

This implies that L is a multiple of the identity and so $dF_y = \frac{h(g)}{h(g_0)} I$. This finishes the proof of Proposition 4.6.

Chapter 5

The Dehn-Nielsen-Baer Theorem

It was stressed that Mostow's Rigidity Theorem does not hold in dimension 2. Thus, nothing has been proved about surfaces in this dissertation yet. However, for surfaces of genus g, the Dehn-Nielsen-Baer theorem is an analog of Corollary 2.5. Recall that this result states that for a manifold M satisfying the hypotheses of Mostow rigidity, we have $\mathrm{Out}(\pi_1(M)) = \mathrm{Isom}(M)$. In the current case, outer automorphisms do not necessarily arise from isometries, but they do arise from homeomorphisms. Namely, we will show that for $g \geq 1$,

$$\mathrm{Mod}^{\pm}(\Sigma_g) = \mathrm{Out}(\pi_1(\Sigma_g)),$$

where $\mathrm{Mod}^{\pm}(\Sigma_g)$ is the generalized mapping class group of Σ_g. This is a remarkable result of algebraic topology, since it relates a purely topological object ($\mathrm{Mod}^{\pm}(\Sigma_g)$) to a purely algebraic object ($\mathrm{Out}(\pi_1(\Sigma_g))$). We have included this result because it can be proved using hyperbolic geometry and quasi-isometries. We follow the approach of Farb and Margalit [7] and we also used [9].

Let us fix some terminology and notation. A surface is a connected orientable 2-manifold of finite type (i.e. its fundamental group is finitely generated). An ***isotopy*** of S is a homotopy $H : S \times [0, 1] \to S$ with the property that $H(\cdot, t)$ is a homeomorphism onto its image for all $t \in [0, 1]$. We denote $\mathrm{Homeo}^0(S)$ the group of homeomorphisms isotopic to the identity map.

Our object of interest will be the ***generalized mapping class group***, which is

$$\mathrm{Mod}^{\pm}(S) = \mathrm{Homeo}(S)/\,\mathrm{Homeo}^0(S).$$

A mapping class $[\varphi] \in \mathrm{Mod}^{\pm}(S)$ does not necessarily induce a unique isomorphism of $\pi_1(S)$. Indeed, if two homeomorphisms φ and φ' of S fixing a basepoint x_0 are isotopic via an isotopy that does not fix x_0, then φ_* and φ'_* will differ by conjugation by the loop $\gamma(t) = H(x_0, t)$. Therefore, a mapping class only induces a well-defined outer automorphism of $\pi_1(S)$.

Theorem 5.1. Dehn-Nielsen-Baer Theorem. *Let $g \geq 1$. The natural map*

$$\begin{array}{ccc} \mathrm{Mod}^{\pm}(\Sigma_g) & \longrightarrow & \mathrm{Out}(\pi_1(\Sigma_g)) \\ \varphi & \longmapsto & [\varphi_*] \end{array}$$

is an isomorphism.

Injectivity of the natural map easily follows from homotopy theory. The universal cover of Σ_g is diffeomorphic to $\mathbb{R}^n$, so Σ_g is a $K(\pi_1(\Sigma_g), 1)$ space. We use again the fact that maps to $K(G, 1)$ spaces inducing the same map on fundamental groups must be homotopic (see p.29). This does not tell us that these maps are isotopic, but it is a fact due to Baer that for closed surfaces other than the closed disk and the closed annulus, any homotopy of homeomorphisms can be improved to an isotopy of homeomorphisms (see [7, 1.4]). This proves injectivity.

For the case $g = 1$, we cannot use hyperbolic geometry (see page 20), but a short direct proof of surjectivity is possible. Since the fundamental group of the torus is abelian, it is clear that $\mathrm{Out}(\pi_1(\Sigma_1)) = \mathrm{Aut}(\mathbb{Z}^2) = \mathrm{GL}(2, \mathbb{Z})$. Furthermore, an element Φ of $\mathrm{Out}(\pi_1(\Sigma_1)) = \mathrm{GL}(2, \mathbb{Z})$ sends the vertices of the square $[0, 1]^2$ to vertices of another parallelogram P of unit area. This readily extends to a map $\varphi : [0, 1]^2 \to P$. This map in turn induces a homeomorphism of Σ_1 that maps to the initial (outer) automorphism Φ. Thus the theorem is proved when $g = 1$. Also notice that this argument shows that $\mathrm{Mod}^{\pm}(\Sigma_1) = \mathrm{GL}(2, \mathbb{Z})$.

A complete proof of the Dehn-Nielsen-Baer Theorem will be given, but we first outline it.

Outline of proof of Dehn-Nielsen-Baer Theorem. Given $[\Phi] \in \mathrm{Out}(\pi_1(\Sigma_g))$, we have to find a homeomorphism φ of Σ_g inducing the outer automorphism $[\Phi]$. The tool that will give rise to φ is nonseparating chains of simple closed curves (see Figure 5.3). Since any essential closed curve corresponds to an element of $\pi_1(\Sigma_g)$, the automorphism Φ acts on such curves. The key of the proof is that Φ sends a nonseparating chain of simple closed curves to another one. This step uses hyperbolic geometry and quasi-isometries. As a consequence of the change of coordinate principle, there exists a map $\varphi \in \mathrm{Homeo}(\Sigma_g)$ having the same effect as Φ on the chain of curves. Then a clever induction argument shows that $[\varphi_*] = [\Phi] \in \mathrm{Out}(\pi_1(\Sigma_g))$.

5.1 Change of coordinates principle

The change of coordinates principle is useful to prove topological statements about curves on surfaces. It generally enables to carry out arguments using some standard picture. For example, given a complicated curve α on some surface S, one can ask whether there exists a curve γ on S with $i(\alpha, \gamma) = 0$. This kind of question is most easily answered if one can notice that there is a homeomorphism mapping α to a simpler curve β.

For us, the change of coordinates principle will give rise to the homeomorphism of Σ_g that we are looking for. Indeed, we will show that any nonseparating chains of curves are homeomorphic (see Figure 5.3 to see our favorite chain of curves). In this section, S will be a closed surface. For simplicity, we will not consider punctured surfaces.

A curve α on S is said to be ***nonseparating*** if $S - \alpha$ is connected.

Proposition 5.2. *For any two nonseparating curves α, β on S, there exists a homeomorphism φ of S such that $\varphi(\alpha) = \beta$.*

Proof. Let S_α (resp. S_β) be the compact surfaces obtained by removing a tubular neighborhood of α (resp. β). Thus S_α and S_β are connected and have the same number of boundary components, punctures and Euler characteric. It follows the classification of surfaces (see [7, Section 1.1]) that S_α and S_β are homeomorphic. This homeomorphism readily extends to a homeomorphism of S sending α to β. □

Before defining chain of curves, another concept is needed. The ***geometric intersection number*** of two transverse curves α, β, denoted $i(\alpha, \beta)$ is the number of intersection points of α, β. The ***algebraic intersection number*** of two transverse curves α, β, denoted $\hat{i}(\alpha, \beta)$, is the sum of indices of intersection points of α and β. An intersection point has index $+1$ if the orientation of the crossing agrees with the orientation of S and -1 otherwise. These quantities can be defined for isotopy classes of curves by taking the minimum over all pairs of representatives.

Definition 5.3. A ***chain*** on S is a sequence of simple closed curves $\alpha_1, \dots, \alpha_r$ with the property that $i(\alpha_i, \alpha_{i+1}) = 1$ for $i = 1, \dots, r-1$ and $i(\alpha_i, \alpha_j) = 0$ whenever $|i-j| \geq 2$. A chain is oriented if the index of the intersection point of α_i and α_{i+1} does not depend on i. A chain is nonseparating is the union $\alpha_1 \cup \ldots \cup \alpha_r$ does not separate S. Chains of isotopy classes of simple closed curves are defined similarly.

The change of coordinate principle for chains of curves is the following.

Proposition 5.4. *Any two nonseparating oriented chains of simple closed curves with the same number of curves are homeomorphic.*

Proof. The proof is by induction on the number of curves. The basis step follows directly from Proposition 5.2.

For the inductive step, let $(\alpha_1, \dots, \alpha_{k+1})$, $(\beta_1, \dots, \beta_{k+1})$ be oriented chains of simple closed curves. Suppose that the chains $(\alpha_1, \dots, \alpha_k)$, $(\beta_1, \dots, \beta_k)$ are homeomorphic and that the homeomorphism agrees with their orientations. Denote by $S_{\alpha,k}$ the surface obtained by removing tubular neighborhoods of $\alpha_1, \dots, \alpha_k$ from S and define $S_{\beta,k}$ similarly. Then $S_{\alpha,k}$ and $S_{\beta,k}$ are homeomorphic and so we can identify them. On $S_{\alpha,k}$, the curves α_{k+1}, β_{k+1} become nonseparating arcs that connect the same pair of boundary components. Thus $S_{\alpha,k+1}$ and $S_{\beta,k+1}$ have the same number of boundary components and the same Euler characteristic. By the classification of surfaces, they are homeomorphic. Moreover, the chains are oriented and homeomorphisms either preserve of reverse the orientation of the whole surface, so that this homeomorphics respects the orientation of the chains. □

The chain of curves on Σ_g that will interest us all have even length $2g$. We will need to know that such chains of curves always are nonseparating.

Proposition 5.5. *Any chain of curves of even length is nonseparating.*

Proof. For simplicity assume that our surface S has no boundary components. Let $\alpha_1,\dots,\alpha_k$ be a chain of curves on S such that $\hat{i}(\alpha_i, \alpha_{i+1}) = 1$. Let U be a tubular neighborhood of the collection of curves.

If there is only one curve, cutting U along α_1 yields a space homeomorphic to a disjoint union of two strips (a strip is topologically $S^1 \times [0,1]$) as shown on the left of Figure 5.1. This is the base step of the induction.

If $k = 2$, the fact that $i(\alpha_1, \alpha_2) = 1$ implies that α_2 joins the two boundary components of the space obtained in the base step. Thus, cutting along α_2

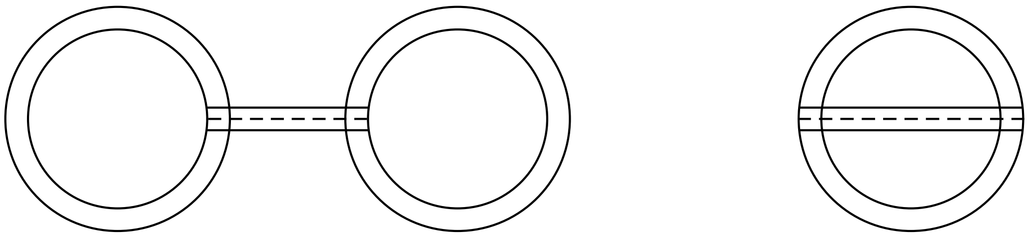


Figure 5.1: Induction by cutting strips

yields a single strip. Thus $U - \{\alpha_1, \alpha_2\}$ is connected and the same holds for $S - \{\alpha_1, \alpha_2\}$. Cutting the resulting surface along α_3 splits the single strip into two strips, as shown on the right hand side of Figure 5.1, and thus we are back to the first inductive step.

Continuing the induction process shows that $U - \{\alpha_1, \ldots, \alpha_k\}$ is connected whenever k is even and disconnected whenever k is odd. This implies that $S - \{\alpha_1, \ldots, \alpha_k\}$ is connected when k is even. In other words, any chain of curves of even length is nonseparating. Finally note that there may be chains of odd length that are nonseparating. □

Free homotopy classes of curves on hyperbolic surfaces are easier to handle because they admit unique geodesic representatives. The lifts of these representatives will be used throughout the proof of the Dehn-Nielsen-Baer Theorem.

Proposition 5.6. *Let S be a closed hyperbolic surface. Then every essential closed curve γ on S is freely homotopic to a unique closed geodesic α. Moreover, if γ is simple then α is simple.*

Proof. The geodesic α is the projection of the axis corresponding to γ. Since images of homotopies are compact, the lifts of two homotopic geodesics to $\mathbb{H}^2$ stay at bounded distance, which forces them to be equal. □

5.2 Quasi-isometries

Quasi-isometries were introduced in Definition 2.7. A finitely generated group G can be made into a metric space via its Cayley graph and the word metric with respect to a set of generators. Choose a symmetric set S of generators for G. The Cayley graph $\mathcal{C}(G, S)$ of G with respect to S has G as set of vertices and two vertices g, h are connected by an edge if there is an $s \in S$ such that $g = sh$. The word metric $d_S(\cdot, \cdot)$ on G is defined as the graph metric on $\mathcal{C}(G, S)$ (edges have length 1). In other words, $d_S(g, h)$ is the reduced word length of gh^{-1}. It is an easy exercise to show that word metrics with respect to different *finite* generating sets are quasi-isometric.

In the case of the fundamental group of a hyperbolic surface, we can give $\pi_1(S)$ a hyperbolic metric as follows. Fix a covering projection with basepoints $p : (\mathbb{H}^2, y_0) \to (S, x_0)$ and define $d_H(\gamma, \delta)$ to be the hyperbolic distance between $\gamma.y_0$ and $\delta.y_0$. Observe that a choice of a different covering map yields a metric quasi-isometric to d_H.

Word metrics and hyperbolic metrics on $G = \pi_1(\Sigma_g)$ are quasi-isometric, which will allow us to switch freely from one to another. Since word metrics are quasi-isometric, we can choose the generating set $S = (a_1, b_1, \ldots, a_{2g}, b_{2g})$. Proving that d_S and d_H are quasi-isometric is an easy exercise. As a hint, one bound comes from the maximum hyperbolic distance between generators in S and the other can be expressed using the injectivity radius of a fundamental domain for Σ_g.

As it was discussed in Section 1.1, the fundamental group of a compact hyperbolic manifolds only contains hyperbolic elements. Thus, throughout this chapter, elements of the fundamental group of Σ_g, $g \geq 2$, will be seen as axes of the underlying hyperbolic deck transformations (often without mention).

Proposition 5.7. *Let $\Psi : G \to G$ be a group automorphism. Then Ψ is a quasi-isometry of G.*

Proof. Left as an exercise. □

For a hyperbolic surface S, elements γ, δ of $\pi_1(S)$ are said to be ***linked at infinity*** if the axes in $\mathbb{H}^2$ corresponding to γ and δ intersect in one point. This is the very property that connects topology to hyperbolic geometry in the proof of the Dehn-Nielsen-Baer theorem. The next lemma is the essential geometric tool of the proof.

Lemma 5.8. *Let Φ be an automorphism of $\pi_1(\Sigma_g)$. Then γ and δ are linked at infinity if and only if and only if $\Phi(\gamma)$ and $\Phi(\delta)$ are linked at infinity.*

Proof. Since Φ is an automorphism, it suffices to show that $\Phi(\gamma)$ and $\Phi(\delta)$

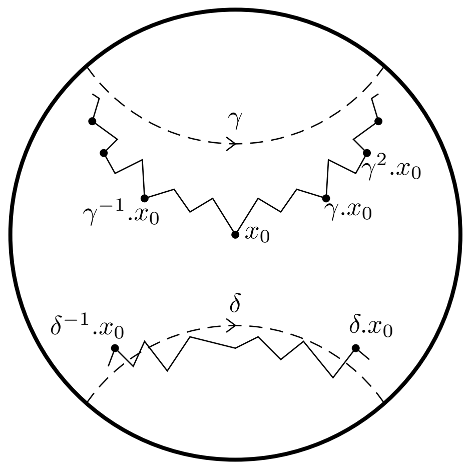


Figure 5.2: Piecewise polygonal paths $P_{\mathcal{O}}$ and $P_{\mathcal{O}'}$

are not linked at infinity whenever γ and δ are not linked at infinity. By Proposition 5.7, Φ is a (K, ε)-quasi-isometry for some $K > 1$, $\varepsilon > 0$.

We are going to use the graph of the universal cover. Fix a covering projection p_0 and a basepoint y_0 in $\mathbb{H}^2$. Choose a fundamental hyperbolic $4g$-gon that has y_0 as a vertex and corresponding to the usual polygonal gluing. This $4g$-gon naturally gives rise to a generating set $(a_1, b_1, \dots, a_g, b_g)$ for $\pi(\Sigma_g)$. The edges of resulting tiling of $\mathbb{H}^2$ form a graph that we simply call the graph of the universal cover. Let $D > 0$ be the maximal hyperbolic length of an edge in that graph. Define

$$\mathcal{O} = \{\gamma^k y_0 : k \in \mathbb{Z}\}$$

and

$$\mathcal{O}' = \{\delta^{kN} y_0 : k \in \mathbb{Z} - \{0\}\}.$$

Choose N so large that $\mathcal{O}$ and $\mathcal{O}'$ are at distance at least $M := K^2 D + 2\varepsilon K + 1$. First connect the points in $\mathcal{O}$ by edges in the graph of the universal cover so that the resulting piecewise geodesic path $P_{\mathcal{O}}$ stays at distance M from $\mathcal{O}'$. Then connect points in $\mathcal{O}'$ in the same way so that the path $P_{\mathcal{O}'}$ stays at distance M away from $P_{\mathcal{O}}$. This construction fails if γ and δ are linked at infinity.

Note that the length of an edge in $\Phi(P_{\mathcal{O}})$ or $\Phi(P_{\mathcal{O}'})$ is at most $KD + \varepsilon$. Assume by contradiction that $\Phi(\gamma)$ and $\Phi(\delta)$ are linked at infinity. Then there exist vertices $y \in P_{\mathcal{O}}$, $y' \in P_{\mathcal{O}'}$ such that

$$KD + \varepsilon \geq d_H(\Phi(y), \Phi(y')).$$

But on the other hand,

$$\begin{aligned} d_H(\Phi(y), \Phi(y')) &\geq \frac{1}{K}M - \varepsilon \\ &\geq KD + \varepsilon + \frac{1}{K}, \end{aligned}$$

which is a contradiction. □

Another relation between elements of $\pi_1(S)$ will be used. Let α, β, $\gamma \in \pi_1(S)$, such that α and β do not share an axis with γ. Then α and β are said to be on the ***same side*** as γ if there is a geodesic δ crossing α, β but not γ.

Corollary 5.9. *Let S be a hyperbolic surface and let α, β, $\gamma \in \pi_1(S)$. Then α, β are on the same side of γ if and only if $\Phi(\alpha)$, $\Phi(\beta)$ are on the same side of $\Phi(\gamma)$.*

Proof. Assume that α, β are on the same side of γ. According to the above definition α and β can be connected by a geodesic δ that does not cross γ. Then apply Lemma 5.8. □

5.3 End of proof

Let $(c_1, c_2, \ldots, c_{2g})$ be a chain of curves on Σ_g as on Figure 5.3 and let Φ be an automorphism of $\pi_1(\Sigma_g)$. We now show that $(\Phi(c_1), \ldots, \Phi(c_{2g}))$ is also a chain of curves on Σ_g. The proof breaks down in four parts.

1. Φ sends simple closed curves to simple closed curves.

2. Φ preserves intersection number 0.

3. Φ preserves intersection number 1.

4. The sign of $\hat{i}(\Phi(c_i), \Phi(c_{i+i}))$ does not depend on the index i.

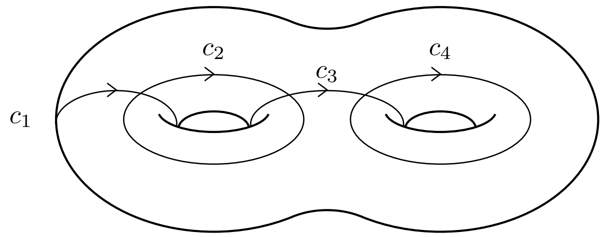


Figure 5.3: Example of a chain of curves on Σ_2

Proof of 1. Let c be an isotopy class of simple closed curve and α be its unique geodesic representative. By Proposition 5.6, α is also simple and thus has minimal period. Since an automorphism sends primitive elements to primitive elements, it follows that the geodesic representative β of $\Phi(c)$ also has minimal period.

There remains to check that β does not intersect itself transversely. We will use implicitly the fact that distinct lifts of α are in one-to-one correspondence with elements conjugate to α in $\pi_1(\Sigma_g)$ (see [9, Prop. 2.5]). Since α is simple, no two of its lifts intersect in $\mathbb{H}^2$. By Lemma 5.8, the image under Φ of two lifts of α do not intersect. Therefore, no two elements in the conjugacy class of $\Phi(c)$ are linked at infinity and so β is simple.

Proof of 2. This relies on the following easy observation. Two isotopy classes of curves have intersection number 0 if and only if no two lifts to $\mathbb{H}^2$ intersect. Since Φ preserves linking at infinity, it preserves intersection number 0.

Proof of 3. This step is more involved and requires a good understanding of the set of lifts of an isotopy class of curves. Let a, b be conjugacy classes of curves and fix a representative α of a. Then $i(a, b) = 1$ if and only if the set of representatives for b linked at infinity with α is $\{\alpha^k \beta \alpha^{-k} : k \in \mathbb{Z}\}$ for some representative β of b linked at infinity with α.

The direction from right to left is obvious. Conversely, since $i(a, b) = 1$, there exists a representative β of b linked at infinity with α. We have to show that any representative of b linked at infinity with a is of the form $\alpha^k \beta \alpha^{-k}$ (the other inclusion is obvious). Let $z = \alpha \cap \beta$ and assume there is a representative β' of b that crosses α strictly between $\alpha^k z$ and $\alpha^{k+1} z$, for

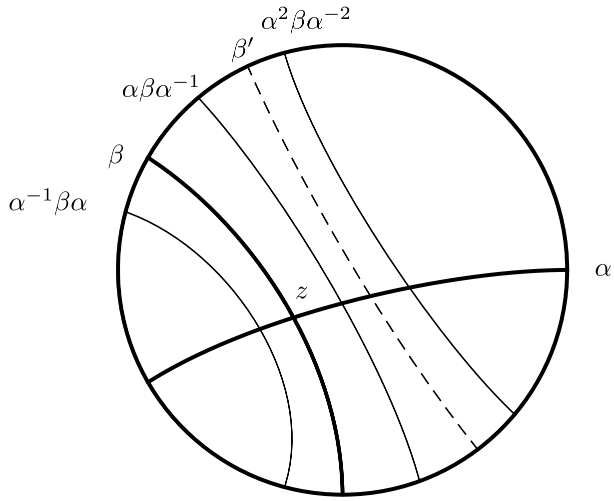


Figure 5.4: Proof of fact 3.

some $k \in \mathbb{Z}$ (see Figure 5.4). By Fact 1, since b is simple, no two lifts of b intersect and so β' is between $\alpha^k\beta\alpha^{-k}$ and $\alpha^{k+1}\beta\alpha^{-(k+1)}$. But this implies that $i(a,b) \geq 2$, which is absurd.

Proof of 4. Let c_1, c_2, c_3 be isotopy classes of curves with $\hat{i}(c_1,c_2) = \hat{i}(c_2,c_3) = \pm 1$ and $i(c_1,c_3) = 0$. Let γ_1, γ_2, γ_3 be the corresponding axes in $\mathbb{H}^2$. The orientation of these curves implies that the lines $\gamma_1\gamma_2\gamma_1^{-1}$ and $\gamma_3^{-1}\gamma_2\gamma_3$ lie on the same side of γ_2. Since Φ preserves the "same side" relation, the images of the c_i's have the right orientation, that is, $\hat{i}(\Phi(c_1),\Phi(c_2)) = \hat{i}(\Phi(c_2),\Phi(c_3)) = \pm 1$.

Now everything is ready to finish the proof of the theorem.

Proof of the Dehn-Nielsen-Baer theorem. Let $[\Phi] \in \mathrm{Out}(\pi_1(\Sigma_g))$ and fix a basepoint $x_0 \in \Sigma_g$. Let $(c_1,\ldots,c_{2g})$ be a chain of curves as above and let $(\gamma_1,\ldots,\gamma_{2g})$ be representatives in $\pi_1(\Sigma_g,x_0)$ that only cross each other at x_0. Notice that the γ_i's generate $\pi_1(\Sigma_g,x_0)$. Since $(\Phi(c_1),\ldots,\Phi(c_{2g}))$ is also a chain of curves, by the change of coordinates principle, there exists a homeomorphism φ such that $\varphi_*(c_i) = \Phi(c_i)$. We can require φ to fix x_0. Thus, $\varphi_*^{-1} \circ \Phi$ is an automorphism of $\pi_1(\Sigma_g)$ fixing conjugacy classes of generators.

It now remains to see that $\varphi_*^{-1} \circ \Phi$ is an inner automorphism. Let I_β denote conjugation by β. We have to find an element $\alpha \in \pi_1(\Sigma_g)$ such that

$$I_\alpha \circ \varphi_*^{-1} \circ \Phi(\gamma_i) = \gamma_i \quad \text{for all } i.$$

Since $\varphi_*^{-1} \circ \Phi(c_1) = c_1$, there exists α such that $I_\alpha \circ \varphi_*^{-1} \circ \Phi(\gamma_1) = \gamma_1$. For γ_2, use the fact proved in the above proof of 3 to see that $I_\alpha \circ \varphi_*^{-1} \circ \Phi(\gamma_2) = \gamma_1^{-k}\gamma_2\gamma_1^{-k}$ for some $k \in \mathbb{Z}$. Therefore, $I_{\gamma_1^k\alpha} \circ \varphi_*^{-1} \circ \Phi(\gamma_j) = \gamma_j$ for $j = 1, 2$.

In fact, the conjugation $I_{\gamma_1^k\alpha}$ satisfies our purpose, which is a little surprising at first sight. That is, we now prove by induction that $I_{\gamma_1^k\alpha} \circ \varphi_*^{-1} \circ \Phi(\gamma_i) = \gamma_i$ for $i = 1, 2, \ldots, 2g$. Since $i(c_1, c_3) = 0$, the elements γ_1 and γ_3 are not linked and so it is easy to see that γ_3 lies strictly between $\gamma_2^l\gamma_1\gamma_2^{-l}$ and $\gamma_2^{l+1}\gamma_1\gamma_2^{-(l+1)}$ for some $l \in \mathbb{Z}$. Moreover, using the fact stated in proof of 3, this uniquely determines γ_3. Since $I_{\gamma_1^k\alpha} \circ \varphi_*^{-1} \circ \Phi$ is an automorphism, it preserves linking at infinity and lying between. The fact that it fixes γ_1 and γ_2 implies that is also fixes γ_3. The inductive step for γ_i is similar, so the proof is finished. $\square$

Bibliography

[1] Tim (ed.) Bedford, Michael (ed.) Keane, and Caroline (ed.) Series. *Ergodic theory, symbolic dynamics, and hyperbolic spaces. Lectures given at the workshop "Hyperbolic geometry and ergodic theory", held at the International Centre for Theoretical Physics in Trieste, Italy, 17-28 April, 1989.* Oxford Science Publications. Oxford etc.: Oxford University Press,. xv, 369 p. , 1991.

[2] R. Benedetti and C. Petronio. *Lectures on hyperbolic geometry.* Universitext. Springer, Berlin, 1992.

[3] G. Besson, G. Courtois, and S. Gallot. Entropies et rigidités des espaces localement symétriques de courbure strictement négative. *Geometric And Functional Analysis*, 5(5):731–799, 1995.

[4] G. Besson, G. Courtois, and S. Gallot. Minimal entropy and Mostow's rigidity theorems. *Ergodic theory and dynamical systems*, 16:623–649, 1996.

[5] Marc Bourdon. Quasi-conformal geometry and Mostow rigidity. preprint, 2000.

[6] M.R. Bridson and A. Haefliger. *Metric spaces of non-positive curvature.* Springer Verlag, 1999.

[7] B. Farb and D. Margalit. A primer on mapping class groups. *preprint*, 2009.

[8] Renato Feres. The Minimal Entropy Theorem and Mostow Rigidity. not meant for publication, 1996.

[9] J. Frazier and A. Sanders. The Dehn-Nielsen-Baer Theorem. October 2008.

[10] U. Haagerup and H.J. Munkholm. Simplices of maximal volume in hyperbolic n-space. *Acta Mathematica*, 147(1):1–11, 1981.

[11] A. Hatcher. *Algebraic topology*. Cambridge University Press, 2002.

[12] Y. Katznelson and B. Weiss. A simple proof of some ergodic theorems. *Israel Journal of Mathematics*, 42(4):291–296, 1982.

[13] R. Mañé and S. Levy. *Ergodic theory and differentiable dynamics*. Springer, Berlin, 1987.

[14] A. Manning. Topological entropy for geodesic flows. *The Annals of Mathematics*, 110(3):567–573, 1979.

[15] Dave Witte Morris. Introduction to arithmetic groups, 2001.

[16] G.D. Mostow. Quasi-conformal mappings in n-space and the rigidity of hyperbolic space forms. *Publications Mathématiques de l'IHÉS*, 34(1):53–104, 1968.

[17] H.J. Munkholm. Simplices of maximal volume in hyperbolic space, Gromov's norm, and Gromov's proof of Mostow's rigidity theorem (following Thurston). In *Topology Symposium, Siegen*, volume 788, pages 109–124, 1979.

[18] S.J. Patterson. The limit set of a Fuchsian group. *Acta mathematica*, 136(1):241–273, 1976.

[19] G. Prasad. Strong rigidity of $\mathbb{Q}$-rank 1 lattices. *Inventiones Mathematicae*, 21(4):255–286, 1973.

[20] John G. Ratcliffe. *Foundations of hyperbolic manifolds. 2nd ed.*, volume 149 of *Graduate Texts in Mathematics*. Springer, New York, 2006.

[21] W. Stepanoff. Sur les conditions de l'existence de la différentielle totale. *Matematicheskii Sbornik*, 32(3):511–527, 1925.

[22] W.P. Thurston. *The geometry and topology of three-manifolds*. Princeton University, Princeton, 1979.

[23] W.P. Thurston and S. Levy. *Three-Dimensional Geometry and Topology: Volume 1*. Princeton University Press, 1997.

[24] P. Walters. *An introduction to ergodic theory, volume 79 of Graduate Texts in Mathematics*. Springer-Verlag, New York, 1982.

Index

str(σ), straightening, 49

arithmetic lattice, 22
asymptotic, 15

boundary at infinity, 15
Busemann function, 19

chain of curves, 78
cone topology, 15
conformal, 18

elliptic isometry, 16
entropy, 60
ergodic, 39

fundamental class, 48

generalized mapping class group, 75
geodesic flow, 41
geodesic ray, 15
Gromov norm, 48

horocyclic flow, 41
hyperbolic isometry, 16

ideal simplex, 50
intersection number
    algebraic, 77
    geometric, 77
isotopy, 75

linked at infinity, 80
Liouville measure, 41
locally symmetric space, 61

measure-preserving transformation, 39

nonseparating curve, 77

parabolic isometry, 16

quasi-conformal, 35
quasi-geodesic, 30
quasi-isometry, 29

regular simplex, 57

same side, 82