Department of Mathematics and Statistics
The Univeristy of Melbourne

Mostow's Rigidity Theorem

James Saunderson

Supervisor: A/Prof. Craig Hodgson

November 7, 2008

Abstract

Mostow's Rigidity Theorem is a stunning bridge between the worlds of geometry and topology. It tells us that the geometry of closed hyperbolic n-manifolds, for $n \geq 3$, is completely determined by their fundamental groups. The proofs of this result are many and varied — all of them substantial and involving machinery from a number of areas of mathematics. The bulk of this report is devoted to giving two proofs of Mostow's Theorem: the first following Besson, Courtois, and Gallot, and the second following Gromov and Thurston. Along the way we will also discuss a tiny fraction of the research inspired by Mostow rigidity and the work that generalizes and extends Mostow's original result.

Contents

Acknowledgements Thanks to my supervisor, Craig Hodgson, for being so patient, knowledgeable, and generous with his time, despite me asking the same questions week after week.

Chapter 1

Introduction

Just over 40 years ago, G. D. Mostow proved a remarkable theorem relating the topology and the geometry of a class of manifolds known as hyperbolic manifolds [Mos68].

For any manifold we can associate an algebraic invariant called the fundamental group. This encodes all of the 'essentially different' ways we can walk around a loop in our manifold and end up back where we started. As such this is a very coarse invariant — there seems little reason to believe that the 'loop structure' of a manifold could give us a lot of detailed information about the manifold. But in the case of 'hyperbolic manifolds' of dimension at least three, it turns out that knowing the fundamental group gives us a great deal of information about the manifold. In fact it determines a unique way to put a 'nice' geometric structure on the manifold — a geometric structure where (as long as you are short-sighted) no matter where you stand, and in which direction you look, the manifold looks the same.

This is the basic idea of Mostow's Strong Rigidity Theorem (which we will state precisely in Chapter 3). Mostow's theorem has inspired a large body of research in the 40 years since it was first proved. This is just as much because the techniques used to prove Mostow Rigidity and its generalizations and extensions contain a great deal of varied, interesting, and deep mathematics, as it is because of the independent interest of the results themselves.

Our primary aim is to give a fairly complete and self-contained proof of Mostow's Theorem. There are at least four quite different approaches to proving Mostow's Theorem: Mostow's original approach [Mos68]; the approach of Agard [Aga83] and Tukia [Tuk85], later refined by Ivanov [Iva96]; the proof given by Gromov and Thurston [Thu79]; and the proof of Besson, Courtois, and Gallot [BCG95]. Of these, the first three are old enough that quite detailed expositions of each are available in various books. For example one can consult Mostow's book [Mos73] for an account of his approach, an account of the approach of Agard, Tukia, and Ivanov is given in [MT98], and the Gromov-Thurston proof is detailed in [Rat94] and [BP92].

On the other hand, the proof of Besson, Courtois, and Gallot is quite recent, dating from 1995. The only expositions of this material in the literature are the original paper [BCG95] (in French), a survey article by the authors [BCG96], and a brief outline in Pansu's survey [Pan97] (also in French). There does not seem to be detailed proof of Mostow rigidity following Besson, Courtois, and Gallot in one place in the literature, so we provide one here.

In Chapter 2 we provide background on hyperbolic geometry and hyperbolic manifolds, an a smattering of topics spanning algebraic topology and measure theory. To avoid over-burdening the reader, the background material is brief, but an effort has been made to introduce the less standard topics that will arise in the remainder of the report.

In Chapter 3 we state Mostow's Rigidity Theorem in two forms and prove their equivalence. We then go on to investigate the non-rigidity that can result when we violate the hypotheses of the theorem by briefly discussing hyperbolic surfaces and lens spaces.

Chapter 4 marks the start of our proof of Mostow Rigidity. In this chapter we prove a result that is the common first step for all the known proofs of Mostow's Theorem. This first step involves showing that any isomorphism between the fundamental groups of two closed hyperbolic manifolds gives rise to a homeomorphism of the boundary of hyperbolic space that satisfies a certain equivariance condition. The standard references typically prove this by first constructing a homotopy equivalence out of the isomorphism between fundamental groups. Our argument is a little different — we avoid this step, instead working directly with the isomorphism between fundamental groups.

In Chapter 5, building on the results of Chapter 4, we give a detailed account of Besson, Courtois, and Gallot's proof of Mostow Rigidity. At the heart of this proof is a construction called the 'barycentric extension', which extends a homeomorphism of the boundary of hyperbolic space to a smooth map with remarkable properties defined on all of hyperbolic space. While much of this material is available in [BCG95], [BCG96], and [Pan97], we reformulate some of the arguments and fill in a number of details that are omitted from the original sources.

In Chapter 6 we sketch the Gromov-Thurston proof of Mostow's Theorem. We do so because the techniques (such as the Gromov norm and measure homology) developed in the course of the proof are quite remarkable. We avoid giving full details in this chapter. The level of detail in our exposition is somewhere between that of the original presentation in [Thu79] and the highly detailed accounts in [BP92] and [Rat94]. The only slightly original contribution in this chapter is our method of constructing the straightening map in Section 6.5.1 using some of the constructions from Chapter 5.

Finally, near the end of each of Chapters 4, 5, and 6 we mention generalizations and extensions of, and similar problems to, Mostow Rigidity. These typically brief sections aim to give the reader a sense of the landscapes opened up by Mostow's original work. We mention the problem of determining the quasi-isometric rigidty of lattices in certain Lie groups (Section 4.5), the more general results of Besson, Courtois, and Gallot that give rise to their proof of Mostow Rigidity as a special case (Section 5.6), and a generalization of Mostow Rigidity due to Thurston (Section 6.8).

Chapter 2

Background

This chapter provides a quick tour of some of the knowledge that will be assumed later in this report. Mostly it is a collection of definitions and facts about hyperbolic geometry, hyperbolic manifolds, and measure theory with a brief detour into the notion of geodesic convexity. Our primary aim, in this chapter, is to fix the notation and terminology that we will use later on.

This chapter quite deliberately falls well short of covering all of the mathematics we will make use of in what follows. In particular, we assume the reader is familiar with some basic facts about smooth manifolds and Riemannian manifolds. Good references for the material on either of these topics are the two books by Lee: [Lee03] and [Lee97]. We also assume familiarity with the basic notions of singular homology. A good reference for this material is Hatcher's book [Hat02].

2.1 Hyperbolic geometry

In this section we introduce n-dimensional hyperbolic space. We do so by describing two models of hyperbolic space — the upper half-space model and the Poincaré model — both of which are simply connected subsets of $\mathbb{R}^n$ together with a Riemannian metric that has constant sectional curvature -1. There are other useful models for hyperbolic space, such as the hyperboloid model and the Klein model (see [CFKP97]), but as we will not make any use of them in what follows, we will not describe them here.

2.1.1 Models of hyperbolic space

The half-space model. Let $\mathcal{H}^n = \{(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n : x_n > 0\}$. If we equip $\mathcal{H}^n$ with the Riemannian metric $ds^2 = \frac{1}{x_n^2} \sum_{i=1}^n dx_i^2$, the resulting Riemannian manifold is n-dimensional hyperbolic space, which we denote by $\mathbb{H}^n$.

The half-space model has a natural boundary $\partial\mathcal{H}^n = \{(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n : x_n = 0\} \cup \{\infty\}$ where ∞ is the point at infinity that compactifies $\mathbb{R}^n$. The geodesics in $\mathcal{H}^n$ are (segments of) circular arcs and lines that are orthogonal to $\partial\mathcal{H}^n$. The geodesic hyperplanes in $\mathcal{H}^n$ are $(n-1)$-spheres and $(n-1)$-planes orthogonal to $\partial\mathcal{H}^n$.

Since the metric at any point is just a rescaled version of the Euclidean metric, it is clear that

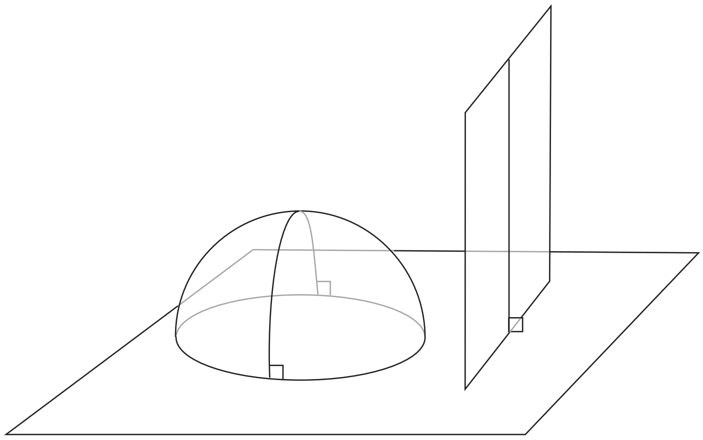


Figure 2.1: Geodesics and hyperplanes in the upper half-space model.

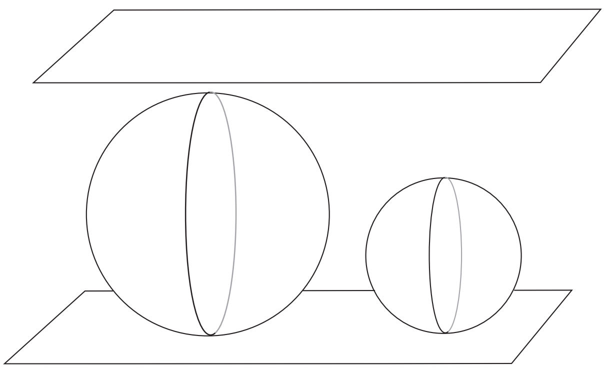


Figure 2.2: Horospheres in the upper half-space model.

the half-space model is conformal. That is, the hyperbolic and Euclidean angles between tangent vectors at any point must coincide.

The group of isometries of $\mathcal{H}^n$ is the group generated by inversions in hyperbolic hyperplanes (which we call 'reflections').

If $\xi \in \partial\mathcal{H}^n \setminus \{\infty\}$ then a *horosphere passing through* ξ is a Euclidean $(n-1)$-sphere tangent to $\partial\mathcal{H}^n$ at ξ. A horosphere in $\mathcal{H}^n$ passing through ∞ is a horizontal Euclidean $(n-1)$-plane, which can be thought of as a 'sphere' 'tangent' to $\partial\mathcal{H}^n$ at the point at infinity. Notice that the Riemannian metric, when restricted to a horosphere passing through ∞, is just the Euclidean metric scaled by a constant. Since all horospheres are isometric to a horosphere passing through ∞, the restriction of the Riemannian metric to any horosphere is a scaled version of the Euclidean metric.

The Poincaré ball model. Let $B^n = \{(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n : \sum_{i=1}^n x_i^2 < 1\}$ be the open unit ball in $\mathbb{R}^n$. Equip B^n with the Riemannian metric

$$ds^2 = 4\frac{\sum_{i=1}^n dx_i^2}{(1 - \sum_{i=1}^n x_i^2)^2}.$$

It turns out that the resulting Riemannian manifold is isometric to $\mathcal{H}^n$. The isometry between B^n and $\mathcal{H}^n$ is given by inversion in the $(n-1)$-sphere of radius $\sqrt{2}$ centered at $(0, 0, \ldots, 0, -1) \in \mathbb{R}^n$.

The boundary ∂B^n of the Poincaré ball model is just the unit sphere in $\mathbb{R}^n$. Note that this is the image of $\partial \mathcal{H}^n$ under the inversion described in the previous paragraph.

Since the Poincaré ball and the upper half-space model are related by an inversion, the Poincaré ball model is conformal. Furthermore, the geodesics, isometries, and horospheres have similar descriptions in both models. Specifically, geodesics in B^n are segments of circular arcs and lines orthogonal to ∂B^n. Hyperplanes are $(n-1)$-spheres and $(n-1)$-planes orthogonal to ∂B^n. The isometries are generated by reflections in hyperplanes. Horospheres in B^n are Euclidean $(n-1)$-spheres tangent to ∂B^n.

Notation. Throughout, we will use the notation $\mathbb{H}^n$ to denote n-dimensional hyperbolic space (independent of any particular model). When we are speficially referring to a particular model of $\mathbb{H}^n$ we will describe the underlying set as either $\mathcal{H}^n$ for the upper half-space in $\mathbb{R}^n$ and B^n for the unit ball in $\mathbb{R}^n$.

We will denote the abstract group of isometries of $\mathbb{H}^n$ by $\mathrm{Isom}(\mathbb{H}^n)$ and the subgroup of orientation preserving isometries by $\mathrm{Isom}_+(\mathbb{H}^n)$. The Riemannian metric on $\mathbb{H}^n$ will be denoted $\langle \cdot, \cdot \rangle$. This symbol will occasionally be used for other purposes. When this occurs, the intended meaning will be explicitly clarified.

2.1.2 The boundary 'at infinity' of hyperbolic space

In both of the models of $\mathbb{H}^n$ we introduced in Section 2.1.1, there was a natural 'boundary' that, when added to $\mathbb{H}^n$, gave a compact topological space. It turns out that this boundary can be defined intrinsically.

Definition 2.1. If X is a metric space and $A, B \subseteq X$ then the *Hausdorff distance* between A and B is

$$d_{\mathcal{H}}(A, B) = \inf_{R \geq 0} \{A \subseteq \mathcal{N}_R(B) \text{ and } B \subseteq \mathcal{N}_R(A)\}$$

where if $S \subseteq X$ then $\mathcal{N}_R(S)$ is the R-neighbourhood of S.

It turns out that if X is a metric space then $d_{\mathcal{H}}$ defines a metric on the space of closed sets in X, and a semi-metric on the set of all subsets of X.

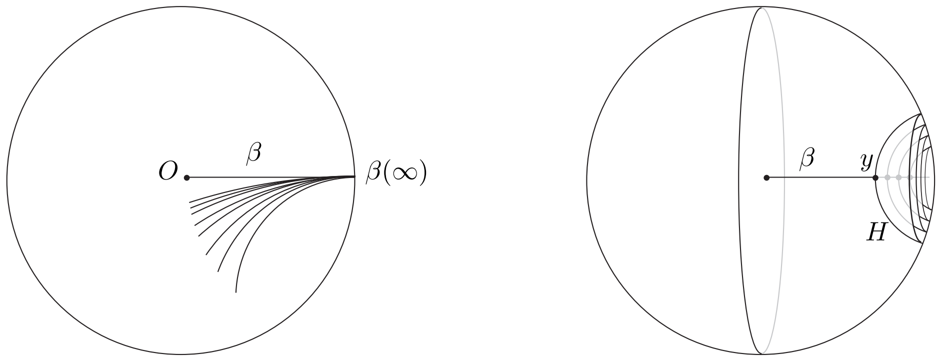
Define an equivalence class on geodesic rays in $\mathbb{H}^n$ as follows. If $\beta_1, \beta_2 : [0, \infty) \to \mathbb{H}^n$ are geodesic rays then $\beta_1 \sim \beta_2$ if and only if the Hausdorff distance between their images is finite. This is clearly an equivalence relation. Define $\partial \mathbb{H}^n$ to be the set of equivalence classes of geodesic rays in $\mathbb{H}^n$. Denote the equivalence class containing the geodesic ray β by $[\beta]$.

From this definition it is easy to see how isometries should act on $\partial \mathbb{H}^n$ and hence on $\overline{\mathbb{H}}^n = \mathbb{H}^n \cup \partial \mathbb{H}^n$. If $\varphi \in \mathrm{Isom}(\mathbb{H}^n)$ and β is a geodesic ray then $\gamma \circ \beta$ is also a geodesic ray so we define

$$\varphi([\beta]) = [\gamma \circ \beta].$$

There are many other ways to describe $\partial \mathbb{H}^n$. In the following two paragraphs we describe two such ways that will be useful later.

In the Poincaré model, all the geodesic rays have a well-defined 'endpoint'. If $\beta : [0, \infty) \to \mathbb{H}^n$ is a geodesic ray then its endpoint is $\lim_{t \to \infty} \beta(t)$ where the limit is taken with respect to the



(a)
An equivalence class $[\beta] \in \partial\mathbb{H}^n$ of geodesic rays.

(b)
A basis of neighbourhoods for $[\beta] \in \partial\mathbb{H}^n$.

Figure 2.3: Intrinsic definition of the boundary of $\mathbb{H}^n$.

usual topology on the unit ball. We denote such an endpoint by $\beta(\infty)$. In any equivalence class of geodesic rays, all the geodesic rays have the same endpoint, so we will often think of points in $\partial\mathbb{H}^n$ in this way.

Finally, given a basepoint $O \in \mathbb{H}^n$ there is a unique geodesic ray emanating from O in each equivalence class of geodesic rays. This can be specified by its initial velocity vector, a unit vector in $T_O\mathbb{H}^n$. Hence we can also identify $\partial\mathbb{H}^n$ with $UT_O\mathbb{H}^n$, the unit sphere in $T_O\mathbb{H}^n$. This will allow us to concretely think of $\partial\mathbb{H}^n$ as a sphere regardless of the model of $\mathbb{H}^n$ in which we are working.

The topology on $\overline{\mathbb{H}}^n$. In the Poincaré model, the topology on $B \cup \partial B$ is just the usual Euclidean topology. We want to give an intrisic description of the topology on $\partial\mathbb{H}^n$ so that $\overline{\mathbb{H}}^n = \mathbb{H}^n \cup \partial\mathbb{H}^n$ is a compact topological space.

We do so by specifying a basis for the topology. Given any point $[\beta] \in \partial\mathbb{H}^n$ (where $\beta : [0, \infty) \to \mathbb{H}^n$ is a geodesic ray), and some point y on β, let H be the hyperplane orthogonal to β and passing through y. Let Q_y^β denote the component of $\overline{\mathbb{H}}^n \setminus H$ containing the point $[\beta] \in \partial\mathbb{H}^n$. Then the Q_y^β are a basis of neighbourhoods for $[\beta]$. Although choosing different representatives from each equivalence class gives a different basis for the topology on $\partial\mathbb{H}^n$ it turns out that the topology itself is independent of the choices made.

2.1.3 Miscellaneous facts about hyperbolic space

In this section we describe a number of facts about hyperbolic space that we will use later on.

Hyperbolic triangles and trigonometry. Hyperbolic triangles are triangles in $\mathbb{H}^n$ with geodesic sides. Remarkably, a hyperbolic triangle with interior angles α, β, and γ has area $\pi - \alpha - \beta - \gamma$ and so the sum of its interior angles is at most π (see [CFKP97, Section 13] for a proof).

There are a number of trigonometric formulae relating the side-lengths and angles of hyperbolic triangles. These can be found, together with a nice unified derivation, in Thurston's book [Thu97, Section 2.4]. We will repeatedly use special cases of one of these formulas, so we will state it here.

Lemma 2.2 (Hyperbolic cosine rule)**.** *If $\alpha, \beta,$ and γ are the interior angles of a hyperbolic triangle and c is the length of the side opposite γ then*

$$\cosh(c) = \frac{\cos(\gamma) + \cos(\alpha)\cos(\beta)}{\sin(\alpha)\sin(\beta)}.$$

In fact, we will be most interested in two special cases. Suppose the triangle has a single 'ideal' vertex (i.e. the vertex is in $\partial\mathbb{H}^n$). Then if its finite side has length c, the angle opposite c is $\gamma = 0$. Hence

$$\cosh(c) = \frac{1 + \cos(\alpha)\cos(\beta)}{\sin(\alpha)\sin(\beta)}. \tag{2.1}$$

Furthermore, if β is a right angle we have

$$\cosh(c) = \frac{1}{\sin(\alpha)}. \tag{2.2}$$

Balls in $\mathbb{H}^n$. If $x \in \mathbb{H}^n$ and $R \geq 0$ then let $B(x; R)$ denote the closed ball of (hyperbolic) radius R centred at x. For any $x \in B^n$ and $R \geq 0$ the hyperbolic ball $B(x; R)$ coincides with a Euclidean ball, although the centre and radius of the Euclidean ball is, in general, different to the centre and radius of the hyperbolic ball [CFKP97]. Since closed Euclidean balls in $\mathbb{R}^n$ are compact, it follows that closed balls are compact in $\mathbb{H}^n$. So by the Hopf-Rinow theorem, $\mathbb{H}^n$ is complete as a metric space.

The volume of a hyperbolic ball of hyperbolic radius R is given by

$$\mathrm{Vol}(B(x; R)) = K_n \int_0^R \sinh^{n-1}(s)\, ds \tag{2.3}$$

where K_n is a constant chosen so that the lowest order behaviour of $\mathrm{Vol}(B(x; R))$ gives the volume of a Euclidean n-ball of radius R. This is obtained by integrating the surface area of a hyperbolic $(n-1)$-sphere. The surface area, in turn, can be deduced from the 'spherical coordinates' on $\mathbb{H}^n$ (see [Rat94, Section 4.7]).

In Chapter 5 we will be interested in the behaviour of the volume of balls for large R. Note that by writing $\sinh(s)$ in terms of e^s it is easy to see that $\mathrm{Vol}(B(x; R)) \sim e^{(n-1)R}$ as $R \to \infty$.

Finally, it follows from (2.3) that in dimension 2 the volume of a disc of hyperbolic radius R is exactly

$$2\pi(\cosh(R) - 1). \tag{2.4}$$

$\mathbb{H}^n$ is uniquely geodesic. Between any pair of distinct points $x, y \in \mathbb{H}^n$ there is a unique geodesic segment. If $x \in \partial\mathbb{H}^n$ and $y \in \mathbb{H}^n$ then there is a unique geodesic ray $\beta : [0, \infty) \to \mathbb{H}^n$ with $\beta(0) = x$ and $\beta(\infty) = y$. Similarly if $x, y \in \partial\mathbb{H}^n$ there is a unique geodesic line $\beta : \mathbb{R} \to \mathbb{H}^n$ with $\beta(-\infty) = x$ and $\beta(\infty) = y$.

2.2 The Hessian and geodesic convexity

2.2.1 The Hessian

Let (M, g) be a Riemannian manifold and let $f : M \to \mathbb{R}$ be a smooth function. Recall that the gradient vector field ∇f is a smooth vector field on M with the defining property that for all

$x \in M$ and all $u \in T_x M$,

$$d_x f(u) = g(\nabla f(x), u).$$

Investigating how ∇f changes as we move in the direction of some $u \in T_x M$ gives a notion of second derivative. Indeed let us define the *Hessian* of f at x by

$$\mathrm{Hess}_x(f)(u, v) = g(D_u \nabla f, v) \tag{2.5}$$

for all $u, v \in T_x M$. (Note that throughout this document we adopt the slightly non-standard notation of D for the Levi-Civita connection associated with the metric g.)

Using the compatibility of the connection with the metric we can rewrite (2.5) as

$$\mathrm{Hess}_x(f)(u, v) = D_U g(\nabla f, V)(x) - g(\nabla f, D_U V)(x) = UV(f)(x) - (D_U V)(f)(x)$$

where U and V are any smooth vector fields with $U(x) = u$ and $V(x) = v$. This gives the usual definition of the Hessian of a smooth real-valued function. Furthermore, this definition makes sense for any connection, whereas (2.5) only made sense for the Levi-Civita connection.

The Hessian is obviously bilinear. If the connection is symmetric (i.e. $[U, V] = D_U V - D_V U$) then the Hessian is also symmetric. To see this, let U and V be smooth vector fields on M, and observe that

$$\begin{aligned}\mathrm{Hess}(f)(U, V) = UV(f) - (D_U V)(f) &= UV(f) - ([U, V] + D_V U)(f) \\ &= UV(f) - UV(f) + VU(f) - (D_V U)(f) \\ &= VU(f) - (D_V U)(f) \\ &= \mathrm{Hess}(f)(V, U).\end{aligned}$$

Since the Levi-Civita connection is symmetric, the Hessian will always be symmetric in cases of interest to us.

2.2.2 Computing the Hessian

We now present two concrete ways to compute the Hessian. First, note that since the Hessian is symmetric and bilinear we can compute it by computing the associated quadratic form. Now suppose $\gamma : (-\epsilon, \epsilon) \to M$ is a geodesic with $\gamma(0) = x$ and $\gamma'(0) = u$. Since γ is a geodesic, if U is a vector field that extends $\frac{d\gamma}{dt}$ then $D_U U = 0$. Hence

$$\mathrm{Hess}_x(f)(U, U) = UU(f)(x) - (D_U U)(f)(x) = UU(f)(x) = \frac{d^2}{dt^2}(f \circ \gamma)(t)\bigg|_{t=0}.$$

Another method of calculation will be prove useful. Let $P(\gamma)_t^0 : T_{\gamma(t)}M \to T_x M$ be parallel transport along γ. Then we can explicitly write the covariant derivative as

$$(D_u \nabla f)(x) = \frac{d}{dt} P(\gamma)_t^0[\nabla f(\gamma(t))]\bigg|_{t=0}.$$

With this in mind, the Hessian can be expressed as

$$\begin{aligned}\mathrm{Hess}_x(f)(u, u) &= \frac{d}{dt} g\left(P(\gamma)_t^0[\nabla f(\gamma(t))], P(\gamma)_t^0[\gamma'(t)]\right)\bigg|_{t=0} \\ &= \frac{d}{dt} g\left(\nabla_{\gamma(t)} f, \gamma'(t)\right)\bigg|_{t=0}.\end{aligned} \tag{2.6}$$

That is, the Hessian measures the infinitessimal change in the angle, as we move along a geodesic, between the gradient vector field and the velocity vector of the geodesic.

2.2.3 Geodesic convexity in $\mathbb{H}^n$

Since $\mathbb{H}^n$ is not a vector space, the usual definition of a convex function does not make sense. But we can use the fact that there is a unique geodesic segment joining any two points in $\mathbb{H}^n$ to make sense of notions such as convex sets, convex hulls, and convex functions.

Definition 2.3. A subset $X \subset \overline{\mathbb{H}}^n$ is *convex* if for any $x, y \in X$ the image of the geodesic segment joining x and y lies in X.

Definition 2.4. The *convex hull* of a collection of points $x_1, x_2, \ldots, x_n \in \overline{\mathbb{H}}^n$ is the smallest (ordered by inclusion) convex set containing $x_1, x_2, \ldots, x_n$.

Definition 2.5. A function $f : \mathbb{H}^n \to \mathbb{R}$ is *convex* if whenever $x, y \in \mathbb{H}^n$ and $\gamma : [a, b] \to \mathbb{H}^n$ is the geodesic segment joining x and y then

$$(f \circ \gamma)((1-t)a + tb) \leq (1-t)(f \circ \gamma)(a) + t(f \circ \gamma)(b) \tag{2.7}$$

for all $t \in (0, 1)$. If the inequality in (2.7) is strict for all $t \in (0, 1)$ then we say that f is *strictly convex.*

Just as in the Euclidean case, the Hessian of a function will provide a useful criterion for determining whether a function is convex. Indeed the next proposition is a straightforward consequence of the fact that given a function $f : \mathbb{H}^n \to \mathbb{R}$, $\mathrm{Hess}_x(f)(u, u)$ is the second derivative of f along the geodesic passing through x with velocity u.

Proposition 2.6. *If $\mathrm{Hess}_x(f)(u, u) \geq 0$ for all $u \in T_x\mathbb{H}^n$ and all $x \in \mathbb{H}^n$ then f is convex. If $\mathrm{Hess}_x(f)(u, u) > 0$ for all $u \in T_x\mathbb{H}^n \setminus \{0\}$ and all $x \in \mathbb{H}^n$ then f is strictly convex.*

Finally, being strictly convex gives us information about the uniqueness of minima.

Lemma 2.7. *If $f : \mathbb{H}^n \to \mathbb{R}$ is strictly convex then there is at most one point $x \in \mathbb{H}^n$ where f takes a minimum value.*

Proof. If not then suppose f takes its minimum value at two distinct points $x, y \in \mathbb{H}^n$. Then by convexity, f must be constant along the geodesic joining x and y. This contradicts the strict convexity of f. □

2.3 Hyperbolic manifolds

Definition 2.8. Suppose M is a smooth n-manifold. A *hyperbolic structure* on M is a family of 'coordinate charts' $\{(\varphi_i, U_i)\}_{i \in I}$ where

1. each U_i is an open subset of M and $\{U_i\}_{i \in I}$ covers M;
2. each φ_i is a diffeomorphism $\varphi_i : U_i \to \varphi_i(U_i)$ onto an open subset $\varphi_i(U_i)$ of $\mathbb{H}^n$;
3. if $U_i \cap U_j \neq \emptyset$ then the corresponding transition map $\varphi_{ij} : \varphi_i(U_i \cap U_j) \to \varphi_j(U_i \cap U_j)$ given by
$$\varphi_{ij}(x) = (\varphi_j \circ \varphi_i^{-1})(x) \quad \text{for all } x \in \varphi_i(U_i \cap U_j)$$
agrees with an element of $\mathrm{Isom}(\mathbb{H}^n)$ on each connected component of $\varphi(U_i \cap U_j)$.

Definition 2.9. If M is a smooth n-manifold equipped with a hyperbolic structure then we say that M is a *hyperbolic n-manifold.*

A hyperbolic manifold can be thought of as a collection of open subsets of $\mathbb{H}^n$ glued together by elements of $\mathrm{Isom}(H^n)$. If all the transition maps are orientation preserving then the resulting hyperbolic manifold will be orientable. The converse is also true: any orientable hyperbolic manifold has a hyperbolic structure where all the transition maps are orientation preserving [BP92, Remark B.1.1].

There are other ways to define 'hyperbolic manifold'. One alternative is to call a Riemannian manifold 'hyperbolic' if its Riemannian metric has constant sectional curvature -1. It is easy to see that any manifold that is hyperbolic according to Definition 2.9 is hyperbolic in this new sense. This is because given a manifold that is hyperbolic according to Definition 2.9, we can define a metric at each point x in the manifold by pulling back the metric on $\mathbb{H}^n$ by a coordinate chart $\varphi_i : U_i \to \varphi_i(U_i)$ with $x \in U_i$. This metric is well-defined precisely because we have insisted that the transition maps agree (locally) with hyperbolic isometries.

Definition 2.10. Two hyperbolic manifolds M and N are *isometric* if there is a diffeomorphism $F : M \to N$ that is an isometry in 'local coordinates'. That is, for all $x \in M$ there is a coordinate chart (φ_i, U_i) for M containing x, and a coordinate chart (ψ_j, V_j) for N containing $F(x)$, such that

$$\psi_j \circ F \circ \varphi_i^{-1} : \varphi_i(U_i \cap F^{-1}(V_j)) \to \psi_j(V_j \cap F(U_i))$$

agrees with an element of $\mathrm{Isom}(\mathbb{H}^n)$ on each connected component of $\varphi_i(U_i \cap F^{-1}(V_j))$.

Again, there is an analogous definition when we define hyperbolic manifolds in terms of Riemannian metrics. Then hyperbolic manfiolds M and N are isometric if they are isometric in the usual sense of Riemannian geometry.

2.3.1 Complete hyperbolic manifolds as quotients of $\mathbb{H}^n$

If we restrict our attention to complete hyperbolic manifolds (i.e. those that are complete as metric spaces), we find that these can always be described as quotients of $\mathbb{H}^n$ by the action of a discrete group of isometries. To state this precisely, we first fix some terminology describing group actions.

Definition 2.11. Suppose Γ is a group of homeomorphisms of a locally compact Hausdorff topological space X. Then Γ acts

1. *freely* on X if there exist $x \in X$ and $\gamma \in \Gamma$ such that if $\gamma \cdot x = x$ then $\gamma = \mathrm{id}_\Gamma$;
2. *properly discontinuously* on X if the set $\{\gamma \in \Gamma : \gamma \cdot K \cap K \neq \emptyset\}$ is finite for all compact $K \subset X$.

It is a standard result that the quotient of a locally compact Hausdorff topological space X by the action of a group of homeomorphisms that acts freely and properly discontinuously on X is again a locally compact Hausdorff topological space. Furthermore, the quotient mapping is a covering map.

It turns out that all complete hyperbolic manifolds can be obtained by this construction.

Theorem 2.12. *If M is a complete connected hyperbolic manifold then there is an action of $\pi_1(M)$ on $\mathbb{H}^n$ such that $\pi_1(M)$ acts freely and properly discontinuously by isometries. Furthermore, M is isometrically diffeomorphic to the quotient manifold $\mathbb{H}^n/\pi_1(M)$.*

The key step in the proof of this is to show that the universal cover of any complete hyperbolic manifold is isometrically diffeomorphic to $\mathbb{H}^n$. With that established we know that π_1 acts on the universal cover of M by covering transformations (which are isometries in this case) in such a way that $M = \tilde{M}/\pi_1(M)$. (For more details see [BP92, Section B.1].)

Notice that the quotient manifold $\mathbb{H}^n/\pi_1(M)$ is invariant under inner automorphisms of $\pi_1(M)$. Hence there is no need to specify a basepoint for the fundamental group in the statement of Theorem 2.12.

We will mostly be interested in closed (i.e. compact and without boundary), connected, oriented hyperbolic manifolds. Since compact metric spaces are complete, the previous theorem also applies to closed hyperbolic manifolds.

2.3.2 Isometries revisited

Since closed hyperbolic manifolds can be described as $\mathbb{H}^n/\Gamma$ where Γ is a discrete subgroup of $\mathrm{Isom}(\mathbb{H}^n)$ acting freely, cocompactly, and properly discontinuously on $\mathbb{H}^n$, we have another way to think about isometries between hyperbolic manifolds. If $F : \mathbb{H}^n/\Gamma_1 \to \mathbb{H}^n/\Gamma_2$ is an isometry then F is certainly a homeomorphism so Γ_1 and Γ_2 must be isomorphic by some $\psi : \Gamma_1 \to \Gamma_2$. Then F lifts to an isometry $\tilde{F} : \mathbb{H}^n \to \mathbb{H}^n$ that is equivariant with respect to the actions of Γ_1 and Γ_2 on $\mathbb{H}^n$. That is

$$\tilde{F} \circ \gamma = \psi(\gamma) \circ \tilde{F}$$

for all $\gamma \in \Gamma_1$. Obviously going back the other way is possible. Any isometry $\tilde{F} : \mathbb{H}^n \to \mathbb{H}^n$ that is equivariant with respect to the actions of Γ_1 and Γ_2 descends to an isometry $F : M \to N$.

This translation of maps between manifolds to equivariant self-maps of $\mathbb{H}^n$ (and back again) will prove very useful.

2.3.3 Hyperbolic manifolds and homotopy equivalences

Let M be a hyperbolic manifold. Since the universal cover of M is $\mathbb{H}^n$, and $\mathbb{H}^n$ is contractible, it follows that all the higher homotopy groups (i.e. $\pi_n(M)$ where $n > 1$) of M are trivial. So if two hyperbolic manifolds have isomorphic fundamental groups then all their homotopy groups are isomorphic. With the help of Whitehead's theorem (see [Hat02, Theorem 4.5]), we will be able to say more.

Theorem 2.13 (Whitehead's theorem)**.** *If a map $f : X \to Y$ between connected CW-complexes induces isomorphisms $f_* : \pi_n(X) \to \pi_n(Y)$ for all n, then f is a homotopy equivalence.*

Since we can represent any closed connected hyperbolic manifold as a CW-complex (by, say, considering such a hyperbolic manifold as a compact polyhedron with faces identified), Whitehead's theorem applies in this case.

Corollary 2.14. *If M and N are closed connected hyperbolic manifolds and $f : M \to N$ induces an isomorphism on fundamental groups then f is a homotopy equivalence.*

With a little extra work we can start with an isomorphism of fundamental groups and construct a homotopy equivalence.

Proposition 2.15. *If M and N are closed hyperbolic manifolds with isomorphic fundamental groups then M and N are homotopy equivalent.*

The idea of the proof is to represent M and N as CW-complexes, lift these to CW complexes in $\mathbb{H}^n$, and then build up a map between the lifted CW complexes that induces the desired isomorphism on fundamental groups. The map is built up by first matching up the 0-skeletons of the lifted CW-complexes, ensuring they satisfy the equivariance condition, then matching up the 1-skeletons, and so on. This matching up process can be done in a consistent way precisely because all the homotopy groups of $\mathbb{H}^n$ are trivial. For a detailed proof see [BP92, Theorem C.5.2].

2.4 The degree of a map

A number of the extensions of Mostow rigidity we consider will be stated in terms of the 'degree' of a smooth map between Riemannian manifolds. This is a generalization of the classical notion of the degree of maps between spheres. Here and elsewhere we assume the reader is familiar with the basics of singular homology.

Suppose M is a closed, oriented, smooth n-manifold. That is, all the transition maps in a smooth structure for M are orientation preserving. In this case, the top-dimensional singular homology (with coefficients in $\mathbb{Z}$) satisfies $H_n(M;\mathbb{Z}) \cong \mathbb{Z}$. Similarly, using coefficients in $\mathbb{R}$ we have $H_n(M;\mathbb{R}) \cong \mathbb{R}$.

Let $[M]$ denote a generator for $H_n(M;\mathbb{Z})$. Then $[M]$ is called a *fundamental class* for M. Given two oriented smooth n-manifolds M and N and a smooth map $f : M \to N$, the *degree* of f is the integer $\deg(f)$ such that

$$f_*([M]) = \deg(f)[N]$$

where $f_* : H_n(M;\mathbb{Z}) \to H_n(N;\mathbb{Z})$ is the map on homology induced by f.

Suppose, instead, we are dealing with homology with coefficients in $\mathbb{R}$ (as will be the case in Chapter 6). Then if $[M]$ generates $H_n(M;\mathbb{R})$ we will call the image of $[M]$ under the inclusion $H_n(M;\mathbb{Z}) \hookrightarrow H_n(M;\mathbb{R})$ a *fundamental class* of M and continue to denote it by $[M]$. With this convention established, if $f : M \to N$ is a smooth map the relationship $f_*([M]) = \deg(f)[N]$ still holds. In this case $f_* : H_n(M;\mathbb{R}) \to H_n(N;\mathbb{R})$ is the map on homology with coefficients in $\mathbb{R}$ induced by f.

2.5 Measure theory

We assume the reader is familiar with the basics of measure theory and integration. In this section we briefly cover some slightly less familiar topics that will be central to certain parts of the discussion later on. Throughout, we will always think of the pair (X, Σ) as a topological space equipped with its Borel σ-algebra. Our reference for this section is [Bau01].

2.5.1 Radon-Nikodym derivatives

Given any two positive measures μ and ν on (X, Σ) we say that μ is *absolutely continuous* with respect to ν if any null set of ν is also a null set of μ.

The Radon-Nikodym derivative gives a concrete way to describe, in terms of integrals, the relationship between two measures, one of which is absolutely continuous with respect to the other.

Theorem 2.16. *If μ and ν are positive, σ-finite measures on (X, Σ) with μ absolutely continuous with respect to ν, then there is a measurable function $f : X \to \mathbb{R}$ such that for all $A \in \Sigma$*

$$\mu(A) = \int_A f \, d\nu.$$

Moreover f is unique in the sense that if there is some other $g : X \to \mathbb{R}$ satisfying this property then f and g agree almost everywhere.

The function f, whose existence is guaranteed by the theorem, is called the *Radon-Nikodym derivative of μ with respect to ν* and is often denoted by the rather suggestive notation $\frac{d\mu}{d\nu}$.

2.5.2 Signed measures

Definition 2.17. A function $\mu : \Sigma \to \mathbb{R} \cup \{-\infty, \infty\}$ is a *signed measure* on (X, Σ) if

1. $\mu(\emptyset) = 0$;
2. $\mu(\cdot)$ does not take both the values ∞ and $-\infty$;
3. for all sequences $(A_i)_{i \in \mathbb{N}} \subset \Sigma$ where the A_i are disjoint, $\mu(\bigcup_{i \in \mathbb{N}} A_i) = \sum_{i \in \mathbb{N}} \mu(A_i)$.

Definition 2.18. The *support* of a positive measure μ on (X, Σ) is the largest closed set $C \subseteq X$ such that whenever $U \subseteq X$ is an open set with $U \cap C \neq \emptyset$ then $\mu(U) > 0$.

It turns out that signed measures can be decomposed into positive measures.

Theorem 2.19 (Hahn-Jordan decomposition)**.** *If μ is a signed measure on a measurable space (X, Σ) then there are unique positive measures μ_+ and μ_- on (X, Σ) such that μ_+ and μ_- have disjoint support and $\mu(A) = \mu_+(A) - \mu_-(A)$ for all $A \in \Sigma$.*

This allows us to define a notion of the 'size' of a signed measure.

Definition 2.20. If μ is a signed measure on (X, Σ), its *total variation* $\|\mu\|$ is

$$\|\mu\| = \mu_+(X) + \mu_-(X).$$

2.5.3 Pushing forward measures

Let (X, Σ_X) and (Y, Σ_Y) be topological spaces equipped with their respective Borel σ-algebras. Let μ be a signed measure on X. If $f : X \to Y$ is a Borel function we can define a measure on Y by

$$f_*[\mu](A) := \mu(f^{-1}[A])$$

for all $A \in \Sigma_Y$. That this construction produces a signed measure follows from the fact that pre-images behave nicely with respect to set operations.

Integration transforms very nicely with respect to pushed-forward measures. Indeed, using the notation from the previous paragraph, if g is integrable with respect to $f_*[\mu]$ then $g \circ f$ is integrable with respect to μ and

$$\int_Y g \, df_*[\mu] = \int_X g \circ f \, d\mu.$$

This is a straightforward consequence of the definition of integration.

Given a locally compact Hausdorff space X, denote the space of signed Borel measures on X with compact support and finite total variation by $\mathcal{M}(X)$. Denote the space of continuous real-valued functions with compact support, equipped with the compact-open topology, by $C_K(X)$. The Riesz Representation theorem tells us that we can identify $\mathcal{M}(X)$ with $C_K(X)^*$ by

$$\mu \mapsto \left(f \mapsto \int_X f \, d\mu\right).$$

Furthermore, the total variation of a measure coincides with the functional norm of the corresponding linear functional. As such we can think of $\mathcal{M}(X)$ as a normed vector space. In this context, pushforwards have particularly nice properties that will play an important role later on.

Lemma 2.21. *If X and Y are locally compact Hausdorff spaces and $f : X \to Y$ is a Borel function then $f_* : \mathcal{M}(X) \to \mathcal{M}(Y)$ is linear and satisfies $\|f_*[\mu]\| \leq \|\mu\|$ for all $\mu \in \mathcal{M}(X)$.*

Proof. Take some $\mu \in \mathcal{M}(X)$ and, using the identification above, think of μ as a bounded linear functional $\Lambda \in C_K(X)^*$. Then due to the way pushforwards and integrals interact, $f_*[\mu]$ corresponds to the linear functional $\Lambda_f \in C_K(Y)^*$ where $\Lambda_f(g) = \Lambda(g \circ f)$. The map $\Lambda \mapsto \Lambda_f$ is obviously linear. To see that it is a contraction observe that

$$\|\Lambda_f\| = \sup_{g \neq 0} \frac{|\Lambda_f(g)|}{\|g\|} = \sup_{g \neq 0} \frac{|\Lambda(g \circ f)|}{\|g\|} \leq \sup_{g \neq 0} \frac{|\Lambda(g \circ f)|}{\|g \circ f\|} \leq \sup_{h \neq 0} \frac{|\Lambda(h)|}{\|h\|} = \|\Lambda\|. \qquad \square$$

Chapter 3

Statement of Mostow Rigidity

3.1 Rigidity theorems

A ‘rigidity theorem’ is, roughly speaking, a theorem where seemingly weak assumptions combine to give very strong conclusions. Many rigidity theorems come in one of the following two flavours.

One flavour is the ‘local’ rigidity theorem. We start with a class of structures (such as Riemannian metrics on a manifold that admits a hyperbolic metric), and some subclass of structures (hyperbolic metrics on that manifold), and ask whether, if we start with a structure in the subclass, we can deform the structure and stay within the subclass of structures. In local rigidity theorems, only trivial deformations are possible. A prototypical example of this style of rigidity theorem is the theorem of Calabi and Weil ([Cal61], [Wei62]) a precursor to Mostow’s Rigidity Theorem.

Another flavour of rigidity theorem is the ‘strong’ rigidity theorem. In this case we start with two objects (such as hyperbolic manifolds) and a ‘weak’ notion of isomorphism between them (such as having isomorhpic fundamental groups). If the type of objects, and the notion of ‘weak’ isomorphism conspire to imply a ‘strong’ notion of isomorphism (such as being isomertic), then we have a ‘strong’ rigidity theorem. Notice that this style of rigidity theorem is global in nature. Mostow’s Ridigity Theorem is the prototypical example of a theorem of this type.

3.2 Mostow’s Rigidity Theorem

In this section we state Mostow’s Theorem in two equivalent, though seemingly different, ways. We then explore examples and non-examples of the rigidity phenomena promised by the theorem, indicating that relaxing the hypotheses causes the theorem to fail.

3.2.1 Statements of the theorem

Theorem 3.1. *Suppose M_1 and M_2 are closed hyperbolic manifolds of dimension $n \geq 3$. If $f : M_1 \to M_2$ is a homotopy equivalence then f is homotopic to an isometry.*

Theorem 3.2. *Suppose $M_1 = \mathbb{H}^n/\Gamma_1$ and $M_2 = \mathbb{H}^n/\Gamma_2$ are closed hyperbolic manifolds of* *dimension $n \geq 3$. If Γ_1 and Γ_2 are isomorphic then they are are actually conjugate in Isom$(\mathbb{H}^n)$.*

Proposition 3.3. *Theorems 3.1 and 3.2 are equivalent.*

Proof. Suppose Theorem 3.1 holds. Then if $M_1 = \mathbb{H}^n/\Gamma_1$ and $M_2 = \mathbb{H}^n/\Gamma_2$ are closed hyperbolic n-manifolds with $n \geq 3$ and Γ_1 and Γ_2 isomorphic then M_1 and M_2 have isomorphic fundamental groups. Then by Proposition 2.15 they are homotopy equivalent. Theorem 3.1 tells us that the homotopy equivalence can be chosen to be an isometry $F : M_1 \to M_2$ that induces the isomorphism between Γ_1 and Γ_2. But then F lifts to an isometry $\tilde{F} : \mathbb{H}^n \to \mathbb{H}^n$ such that $\tilde{F}$ is equivariant with respect to the action of Γ_1 and Γ_2 on $\mathbb{H}^n$. Finally, this is equivalent to saying that $\tilde{F}$ conjugates Γ_1 to Γ_2.

Conversely, suppose Theorem 3.2 holds. Suppose M_1 and M_2 are closed hyperbolic n-manifolds with $n \geq 3$ and $f : M_1 \to M_2$ is a homotopy equivalence. Then since f induces an isomorphism on fundamental groups, Theorem 3.2 tells us that the fundamental groups of M_1 and M_2 are conjugate in Isom$(\mathbb{H}^n)$ (by $\tilde{F} \in$ Isom$(\mathbb{H}^n)$, say). Then

$$\pi_1(M_1) = \tilde{F}\pi_1(M_2)\tilde{F}^{-1}$$

so $\tilde{F}$ descends to an isometry $F : M_1 \to M_2$ that induces the same isomorphism on fundamental groups as f. Hence f is homotopic to F. □

Although Mostow's theorem holds without any assumptions on the orientability of the manifolds involved, our proofs in Chapters 5 and 6 will actually assume that M_1 and M_2 are orientable.

3.2.2 Failure when $n = 2$: Closed surfaces

Mostow's theorem says nothing about what happens for closed orientable hyperbolic manifolds of dimension 2. These manifolds are exactly the closed orientable surfaces of genus $g \geq 2$, which we denote Σ_g. Are these Mostow rigid? Or do there exist many non-isometric hyperbolic structures on Σ_g?

In stark contrast to the situation in higher dimensions, it turns out that there is a $6g - 6$ dimensional space of possible hyperbolic metrics on Σ_g. We will now sketch one way to see this by decomposing Σ_g into hyperbolic 'pairs of pants'. A hyperbolic pair of pants is a genus-0 surface with a hyperbolic metric and three geodesic boundary components. It turns out that given any

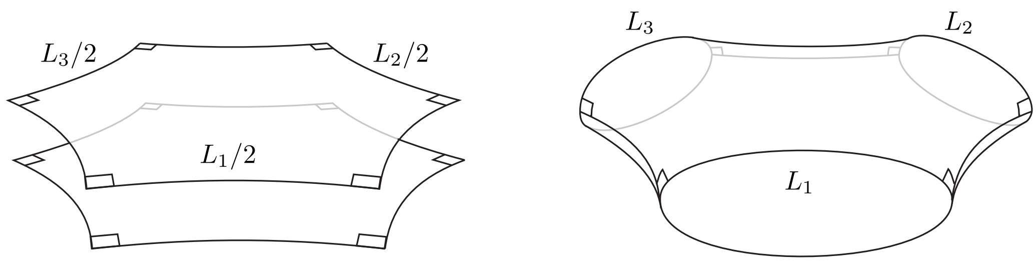


Figure 3.1: Making a pair of hyperbolic pants.

three lengths $L_1, L_2, L_3 > 0$ we can put a unique hyperbolic structure on a pair of pants so that the boundary components have lengths L_1, L_2, and L_3. This is because there is, up to isometry, a unique right-angled hyperbolic hexagon with the side lengths of three non-adjacent sides being $L_1/2$, $L_2/2$, and $L_3/2$ (see [Rat94, Section 3.5]). Gluing two of these together gives a hyperbolic pair of pants. (See Figure 3.1.) So there is a three dimensional space of possible non-isometric hyperbolic structures on a pair of pants. Given any Σ_g with $g \geq 2$ we can decompose the surface

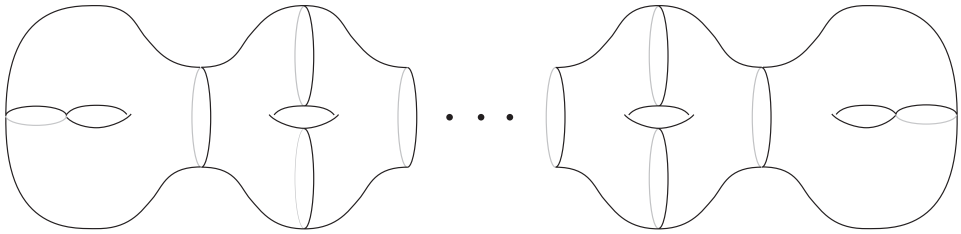


Figure 3.2: Cutting Σ_g into hyperbolic pants.

into $2g-2$ pairs of pants by cutting along $3g-3$ simple closed geodesics. An example of this is shown in Figure 3.2.

Once we have cut our surface into pants, there is a $3(2g-2)$-dimensional space of hyperbolic structures on the disjoint union of the $2g-2$ pairs of pants. Since we want to put a hyperbolic structure on the surface itself, we need to make sure that we can actually glue our pairs of pants back together. To ensure this, we must make sure that the boundary components that need to be re-glued have the same length. This imposes $3g-3$ extra constraints. But along every gluing curve, we can choose to glue the legs of the pants together with a twist of some angle, giving us an extra $3g-3$ degrees of freedom. This quick dimension count suggests there is a $(6g-6)-(3g-3)+(3g-3) = 6g-6$ dimensional space of non-isometric hyperbolic structures on Σ_g. Indeed this is the case.

So hyperbolic surfaces certainly do not satsify Mostow's Rigidity Theorem, as if they did there would have to be a unique hyperbolic structure on each Σ_g where $g \geq 2$.

3.2.3 Failure for spherical manifolds: Lens spaces

Just as in the case of compact hyperbolic n-manifolds, 'spherical' n-manifolds can be described as the quotient of $\mathbb{S}^n$ by the action of a subgroup of $\mathrm{Isom}(\mathbb{S}^n)$ acting freely, properly discontinuously.

In this section we give an example of spherical 3-manifolds that have isomorphic fundamental groups and yet are not homeomorphic (and so are certainly not isometric). This indicates that Mostow's rigidity theorem fails for spherical 3-manifolds.

We can think of $\mathbb{S}^3$ as a subset of $\mathbb{C}^2$ by writing $\mathbb{S}^3 = \{(z_1, z_2) \in \mathbb{C}^2 : |z_1|^2 + |z_2|^2 = 1\}$. Let $\omega \in \mathbb{C}$ be a primitive pth root of unity and let q be relatively prime to p. Then

$$\Gamma_{p,q} = \left\{ \begin{bmatrix} \omega & 0 \\ 0 & \omega^q \end{bmatrix}^k : k = 0, 1, \ldots, p-1 \right\} \cong \mathbb{Z}_p$$

acts on $\mathbb{S}^3$ by left-multiplication. Define, for relatively prime positive integers p, q, the three dimensional *lens space* to be

$$L(p,q) := \mathbb{S}^3/\Gamma_{p,q}.$$

Figure 3.3 shows a way to think of lens spaces in terms of identifying faces on a polyhedron.

From our construction it is clear that for any q, the fundamental group of $L(p,q)$ is $\mathbb{Z}_p$. So for any two lens spaces $L(p,q)$ and $L(p',q')$ to be homotopy equivalent it is necessary that $p = p'$.

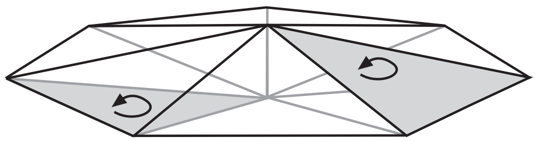


Figure 3.3: The lens space $L(7,2)$ can be described by identifying faces in the polyhedron shown. The polyhedron consists of seven 'wedges', the identification is to glue any bottom face to the top face two 'wedges' along.

In fact the classification of lens spaces up to homotopy equivalence and up to homemorphism are known. A proof of the former is given in [Whi41], a proof of the latter in [Bro60].

Theorem 3.4. *If $L(p,q)$ and $L(p,q')$ are three dimensional lens spaces then they are*

1. *homotopy equivalent if and only if $q \equiv \pm n^2 q' \pmod{p}$ for some $n \in \mathbb{N}$ and*
2. *homeomorphic if and only if $q \equiv \pm q' \pmod{p}$ or $qq' \equiv \pm 1 \pmod{p}$.*

Hence the lens spaces $L(7,1)$ and $L(7,2)$ are homotopy equivalent as $2 \equiv 3^2 \times 1 \pmod{7}$ but are not homeomorphic. So, in particular, Mostow's theorem does not hold for these spherical 3-manifolds.

In a sense, spherical manifolds are bad examples of non-rigidity. This is because they do satisfy a weaker rigidity theorem due to de Rham [dR50]. This states that spherical manifolds that are diffeomorphic are actually isometric.

3.3 More general versions of Mostow's Theorem

Mostow's Theorem (as stated in Section 3.2.1) has been generalized a number of times over the years. Here we will briefly mention the generalizations that are still referred to as 'Mostow Rigidity'.

First, Mostow himself extended the theorem to deal with the case that M_1 and M_2 are 'rank one locally symmetric spaces' (with some exceptions) rather than just hyperbolic manifolds [Mos73]. Locally symmetric spaces are Riemannian manifolds where the 'geodesic reflection' map, that takes $\exp_p(tv)$ to $\exp_p(-tv)$ for sufficiently small t and any v in the tangent space at p, agrees locally with an isometry. The 'rank' of such a space is the largest dimension of a totally geodesic flat submanifold.

It is easy to see that hyperbolic manifolds are locally symmetric. To see that they have rank one, observe that the totally geodesic k-dimensional submanifolds of hyperbolic manifolds are locally isometric to $\mathbb{H}^k$ and so cannot be flat unless $k = 1$.

Prasad extended Mostow's results further by replacing the assumption that the manifolds be compact, with the assumption that they have finite volume [Pra73]. As such, the resulting theorem is sometimes known as Mostow-Prasad Rigidity.

Chapter 4

The first step

While there are many proofs of Mostow's Rigidity Theorem, they all share a similar 'first step' which essentially goes back to Mostow's original proof [Mos68]. Recall that in the statement of the theorem, we start with an isomorphism between fundamental groups of closed hyperbolic n-manifolds. The first step of all the known proofs of Mostow's theorem involves showing that this map can be used to construct a homeomorphism of the sphere 'at infinity' of hyperbolic space.

As such, this chapter is devoted to proving the following theorem.

Theorem 4.1. *Suppose Γ_1 and Γ_2 are subgroups of Isom($\mathbb{H}^n$) such that $M_1 = \mathbb{H}^n/\Gamma_1$ and $M_2 = \mathbb{H}^n/\Gamma_2$ are closed hyperbolic manifolds. If $\psi : \Gamma_1 \to \Gamma_2$ is an isomorphism then there is a homeomorphism $\partial f : \partial\mathbb{H}^n \to \partial\mathbb{H}^n$ such that*

$$\partial f \circ \gamma = \psi(\gamma) \circ \partial f \quad \textit{for all} \quad \gamma \in \Gamma_1.$$

Our proof differs from the standard references ([BP92, Section C.1], [Rat94, Section 11.6], [Thu79, Section 5.9]) in that we use the isomorphism between Γ_1 and Γ_2 directly to construct ∂f, without first constructing a homotopy equivalence between M_1 and M_2.

In Chapter 5 we will prove that ∂f actually must be the restriction to $\partial\mathbb{H}^n$ of an element of Isom($\mathbb{H}^n$), establishing Mostow's Rigidity Theorem.

4.1 Finitely generated groups as geometric objects

Let G be a finitely generated group and let $X \subset G$ be a finite generating set for G. Let X^* denote the set of words in the elements of X and their inverses.

Then we can define a metric on G (depending on X) as follows. If $g \in G$ then define

$$\|g\|_X = \min\{\text{length}(w) : w \in X^* \text{ and } w \underset{G}{=} g\}.$$

Then for any $g, h \in G$ define

$$d_X(g,h) = \|g^{-1}h\|_X.$$

This defines a metric on G. The proof is a routine check and so we omit it. Furthermore it is easy to see from the definition that the metric is left-invariant in the sense that for all $g_1, g_2, h \in G$,

$d_X(hg_1, hg_2) = d_X(g_1, g_2)$. So, as long as we specify a generating set X, we can think of any finitely generated group G as a metric space (G, d_X).

Of course, changing generating sets changes the metric, which is problematic if we want to use the metric space structure to study G itself. Thankfully there is a coarser (but not too coarse) notion of equivalence between metric spaces such that if X and Y are two generating sets for a finitely generated group G, then (G, d_X) and (G, d_Y) are equivalent.

Definition 4.2. Given metric spaces (X, d_X) and (Y, d_Y), a map $f : X \to Y$ is a (λ, ϵ)-*quasi-isometric embedding* if there are constants $\lambda \geq 1$ and $\epsilon \geq 0$ such that for all $x_1, x_2 \in X$

$$\frac{1}{\lambda} d_X(x_1, x_2) - \epsilon \leq d_Y(f(x_1), f(x_2)) \leq \lambda d_X(x_1, x_2) + \epsilon.$$

Furthermore, if

$$Y \subseteq \mathcal{N}_\epsilon(f(X)) := \{y \in Y : d(y, f(x)) \leq \epsilon \text{ for some } x \in X\}$$

then f is ϵ-*quasi-surjective.* If f is both a (λ, ϵ)-quasi-isometric embedding and ϵ-quasi-surjective it is a (λ, ϵ)-*quasi-isometry.* Finally if there is a quasi-isometry $f : X \to Y$ then the metric spaces X and Y are *quasi-isometric.*

It is a simple yet crucial fact that the composition of two quasi-isometries is again a quasi-isometry.

Proposition 4.3. *If X and Y are two finite generating sets for a group G, then the metric spaces (G, d_X) and (G, d_Y) are quasi-isometric.*

Proof. For each $x \in X$ let $w_Y(x) \in Y^*$ denote a fixed minimum length word over Y such that $w_Y(x) \underset{G}{=} x$. Similarly for each $y \in Y$ let $w_X(y) \in X^*$ denote a fixed minimum length word over X such that $w_X(y) \underset{G}{=} y$.

Let $\lambda = \max\{\text{length}(w_X(y)) : y \in Y\}$ and $\mu = \max\{\text{length}(w_Y(x)) : x \in X\}$ and $\eta = \max\{\lambda, \mu\}$. These exist since X and Y are finite.

If $g, h \in G$ and $g^{-1}h \underset{G}{=} x_1 x_2 \cdots x_n$ is a minimum length word representing $g^{-1}h$ over X then $g^{-1}h \underset{G}{=} w_Y(x_1) w_Y(x_2) \cdots w_Y(x_n)$. So

$$d_Y(g, h) = \|g^{-1}h\|_Y \leq \sum_{i=1}^{n} \text{length}(w_Y(x_i)) \leq \eta n = \eta d_X(g, h).$$

Reversing the roles of X and Y gives $d_X(g, h) \leq \eta d_Y(g, h)$ so the identity map $\text{id}_G : (G, d_X) \to (G, d_Y)$ is a quasi-isometric embedding. The identity map is surjective. Hence (G, d_X) and (G, d_Y) are quasi-isometric. $\square$

As a result, if we are only interested in a finitely generated group's metric space structure up to quasi-isometry, we do not need to, and usually will not bother to, specify a particular finite generating set for the group.

Also observe that, with minor alterations to notation, the argument in the proof of Proposition 4.3 shows that if G and H are isomorphic finitely generated groups then for any generating sets X for G and Y for H the isomorphism $\psi : G \to H$ gives rise to a quasi-isometry $\psi : (G, d_X) \to (H, d_Y)$. This fact will play a key role in Section 4.4.

Finally, just as any isometry of metric spaces is invertible and the inverse is also an isometry, an analogous result holds for quasi-isometries.

Definition 4.4. If $f : X \to Y$ is a quasi-isometry then $g : Y \to X$ is a *κ-quasi-inverse* of f if there is some $\kappa > 0$ such that for all $x \in X$, $y \in Y$.

$$d_X((g \circ f)(x), x) \leq \kappa \quad \text{and} \quad d_Y((f \circ g)(y), y) \leq \kappa.$$

Lemma 4.5. *If f is a (λ, ϵ)-quasi-isometry and g is a κ-quasi-inverse of f then g is also a quasi-isometry.*

Proof.

$$\begin{aligned} d_X(g(x), g(y)) &\leq \lambda d_Y((f \circ g)(x), (f \circ g)(y)) + \lambda\epsilon \\ &\leq \lambda d_Y((f \circ g)(x), x) + \lambda d_Y(x, y) + \lambda d_Y(y, (f \circ g)(y)) + \lambda\epsilon \\ &\leq \lambda d_Y(x, y) + \lambda(2\kappa + \epsilon). \end{aligned}$$

Similarly

$$\begin{aligned} d_Y(x, y) &\leq d_Y(x, (f \circ g)(x)) + d_Y((f \circ g)(x), (f \circ g)(y)) + d_Y((f \circ g)(y), y) \\ &\leq \lambda d_X(g(x), g(y)) + (2\kappa + \epsilon). \end{aligned}$$

□

Proposition 4.6. *If $f : X \to Y$ is a (λ, ϵ)-quasi-isometry then f has a quasi-inverse $g : Y \to X$ with the property that $f \circ g \circ f = f$.*

Proof. If $y \in f(X)$ then fix some $x_y \in f^{-1}(\{y\})$ and define $g(y) = x_y$. If $y \notin f(X)$ then there is some $y_0 \in f(X)$ with $d_Y(y, y_0) \leq \epsilon$ since f is ϵ-quasi-surjective. Hence define $g(y) = x_{y_0}$.

It is clear from the definition that $(f \circ g)(y) = y_0$ and so $d_Y((f \circ g)(y), y) \leq \epsilon$. Moreover, $(f \circ g \circ f)(x) = (f(x))_0 = f(x)$ so, since f is a quasi-isometry,

$$d_X((g \circ f)(x), x) \leq \lambda d_Y((f \circ g \circ f)(x), f(x)) + \lambda\epsilon = \lambda\epsilon.$$ □

4.2 δ-hyperbolic spaces

One viewpoint of 'negatively curved spaces' is that they are Riemannian manifolds with negative sectional curvature. Another more general notion of a negatively curved space is that of a δ-hyperbolic space or Gromov hyperbolic space. These are named after M. Gromov whose work has made great use of this general approach to negative curvature. Roughly speaking, a geodesic metric space X is called 'hyperbolic' if all geodesic triangles in X are 'slim'. Let us now make these notions more precise.

Definition 4.7. A *geodesic* in a metric space X is a path $c : [a, b] \to X$ that is an isometric embedding. That is, for all $t_1, t_2 \in [a, b]$,

$$d_X(c(t_1), c(t_2)) = |t_1 - t_2|. \tag{4.1}$$

The main space of interest to us is $\mathbb{H}^n$. Note that the notion of geodesic in Definition 4.7 corresponds to the Riemannian notion of geodesic for $\mathbb{H}^n$. It does *not* correspond to the Riemannian notion of geodesic in general as Riemannian geodesics need not be globally length minimizing. We have already noted that there is a (unique) geodesic joining any two points in $\mathbb{H}^n$. This motivates the following definition.

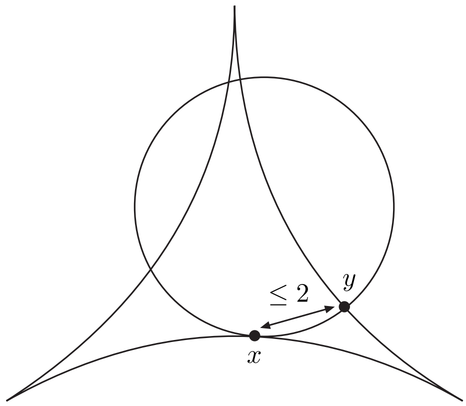


Figure 4.1: Hyperbolic triangles are 2-slim.

Definition 4.8. A metric space X is *geodesic* if for any $x, y \in X$ there is a geodesic joining x and y.

For the sake of convenience, if $x, y \in X$ and there is a geodesic joining x and y, we use the notation $[x, y] \subset X$ to denote the image of some geodesic joining x and y.

Now we are in a position to define a δ-hyperbolic space. The definition we use is attributed to Rips by Gromov [Gro87], and is one of many equivalent definitions that are in use.

Definition 4.9. If X is a geodesic metric space, a geodesic triangle in X is *δ-slim* if each side is contained in the δ-neighbourhood of the union of the other two sides.

Definition 4.10. A geodesic metric space is *δ-hyperbolic* if there is some $\delta > 0$ such that every geodesic triangle in X is δ-slim.

If δ-hyperbolic spaces are meant to exhibit behaviour similar to that of negatively curved Riemannian manifolds, then the simplest negatively curved Riemannian manifold ought to be δ-hyperbolic. Indeed this is the case.

Proposition 4.11. *In $\mathbb{H}^n$ all triangles are 2-slim.*

Proof. Since all geodesic triangles in $\mathbb{H}^n$ lie in some isometric copy of $\mathbb{H}^2$, it suffices to prove this in the case $n = 2$. Recall from Section 2.1.3 that hyperbolic triangles have area at most π and that the area of a hyperbolic disk of radius r is $2\pi(\cosh(r) - 1)$. Any circle inscribed in a hyperbolic triangle must have area less than π so must have radius less than $\cosh^{-1}(\frac{3}{2}) < 1$.

Given a point x on a hyperbolic triangle, construct a circle of radius 1 tangent to the triangle at x. Then the circle intersects another side of the triangle at some y. Since both x and y lie on the circle of radius 1 it follows that $d(x, y) \leq 2$. $\square$

The point of all this is that the slim triangles condition captures some of the large-scale properties of negatively curved Riemannian manifolds but is valid in a much more general setting.

Although we will not need this fact, it is interesting to note that if two geodesic metric spaces (X, d_X) and (Y, d_Y) are quasi-isometric, then X is δ-hyperbolic for some δ if and only if Y is δ'-hyperbolic for some δ' (see [BH99, Theorem III.H.1.9] for a proof). This is further evidence that quasi-isometry is the 'right' notion of equivalence in this context.

4.3 Quasi-geodesics, the Morse lemma, and another way to view $\partial\mathbb{H}^n$

4.3.1 Quasi-geodesics and the Morse lemma

Definition 4.12. If X is a metric space then $c : [a,b] \to X$ is a (λ, ϵ)-*quasi-geodesic* if c is a (λ, ϵ)-quasi-isometric embedding.

Note that the images of geodesics and quasi-geodesics under quasi-isometries are quasi-geodesics.

Quasi-geodesics play an important role in δ-hyperbolic spaces because they are 'not too far away' from being geodesics. This statement about the 'stability of quasi-geodesics' is made precise by the 'Morse Lemma'. A version of this result (in $\mathbb{H}^2$ and restricted to what we will call 'tamed' quasi-geodesics in Section 4.6) was first proved by Morse in [Mor21].

Lemma 4.13 (Morse Lemma)**.** *If X is a δ-hyperbolic space then there is some constant $R = R(\delta, \lambda, \epsilon)$ such that for any (λ, ϵ)-quasi-geodesic $c : [a,b] \to X$,*

$$d_{\mathcal{H}}(c([a,b]), [c(a), c(b)]) \leq R.$$

Since the proof is lengthy enough to interfere with our current line of thought we defer it until Section 4.6.

While the Morse Lemma only tells us about the behaviour of finite quasi-geodesic segments, we can use it to say something about quasi-geodesic rays as well. Since we are really only interested in the case where our δ-hyperbolic space is $\mathbb{H}^n$ we will now focus on that situation.[1]

Lemma 4.14. *If $c : [0,\infty) \to \mathbb{H}^n$ is a (λ, ϵ)-quasi-geodesic then there is a unique geodesic ray $A(c) : [0,\infty) \to \mathbb{H}^n$ that satisfies $A(c)(0) = c(0)$ and is such that the Hausdorff distance between the image of c and the image of $A(c)$ is finite.*

Proof. Let $p = c(0)$. Let $V_p : \mathbb{H}^n \to UT_p\mathbb{H}^n$ denote the 'visual' projection of $\mathbb{H}^n$ onto the unit tangent space at p defined by mapping $x \in \mathbb{H}^n$ to the initial velocity vector of the geodesic joining p and x.

For every $t \in [0,\infty)$ let B_R^t denote the closed ball of radius $R = R(\lambda, \epsilon)$ centred at $c(t)$ and let

$$U = \bigcap_{t\in[0,\infty)} V_p(B_R^t).$$

Note that U is precisely the set of initial velocity vectors of geodesic rays based at p that stay within distance R of $c([0,\infty))$. So it suffices to show that U consists of a single point.

If U contains more than one point then there would be two distinct geodesic rays γ and γ' that remain within a bounded distance of $c([0,\infty))$. But this is impossible as any two such geodesic rays do not remain within a bounded distance of each other.

Now, the collection of closed subsets $\{V_p(B_R^t)\}_t$ of $UT_p\mathbb{H}^n$ satisfy the finite intersection property. This is because for any finite subset $\{t_1, t_2, \ldots t_k\} \subset [0,\infty)$ (with $t_1 < t_2 < \cdots < t_k$), Lemma 4.13 tells us that the geodesic segment $\gamma_k = [c(0), c(t_k)]$ passes through all the balls $B_R^{t_i}$. Hence $\gamma_k'(0)$ is an element of each $V_p(B_R^{t_i})$.

[1]Nevertheless, it is possible to prove most of what follows in the context of δ-hyperbolic spaces.

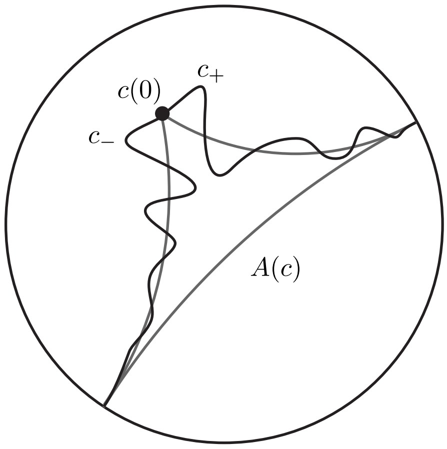


Figure 4.2: There is a geodesic line within a bounded distance of any quasi-geodesic line.

Observe that $UT_p\mathbb{H}^n$ is compact, each B^t_R is compact, and V_p is continuous. So it follows that $\{V_p(B^t_R)\}_t$ is a collection of compact (hence closed) subsets of a compact set that satisfy the finite intersection property. So $\{V_p(B^t_R)\}_t$ has non-empty intersection.

Hence $U = \{u\}$ is a singleton. Thus taking $A(c)$ to be the geodesic ray with $A(c)(0) = p$ and $A(c)'(0) = u$ completes the proof. ☐

Corollary 4.15. *If $c : \mathbb{R} \to \mathbb{H}^n$ is a quasi-geodesic line then there is a unique geodesic line $A(c) : \mathbb{R} \to \mathbb{H}^n$ such that the Hausdorff distance between the images of c and $A(c)$ is bounded.*

Proof. Let $c_+ : [0, \infty) \to \mathbb{H}^n$ and $c_- : [0, \infty) \to \mathbb{H}^n$ be given by $c_+(t) = c(t)$ for $t \in [0, \infty)$ and $c_-(t) = c(-t)$ for $t \in [0, \infty)$. Then let $A(c)$ be the geodesic line which stays within a bounded distance of the geodesic rays $A(c_+)$ and $A(c_-)$. (See Figure 4.2.) ☐

Note that given any quasi-geodesic ray $c : [0, \infty) \to \mathbb{H}^n$ (geodesic line $c : \mathbb{R} \to \mathbb{H}^n$) we will continue, throughout this chapter, to use the notation $A(c) : [0, \infty) \to \mathbb{H}^n$ ($A(c) : \mathbb{R} \to \mathbb{H}^n$) for the unique geodesic ray (geodesic line) that stays within a bounded distance of c.

4.3.2 $\partial\mathbb{H}^n$ from another perspective

Recall that in Section 2.1.2 we defined $\partial\mathbb{H}^n$, the boundary of $\mathbb{H}^n$ 'at infinity', to be the set of equivalence classes of geodesic rays in $\mathbb{H}^n$ where geodesic rays η_1, η_2 are equivalent if the Hausdorff distance between their images is finite.

Similarly we say that quasi-geodesic rays $c_1, c_2 : [0, \infty) \to \mathbb{H}^n$ are equivalent if the Hausdorff distance between their images is finite. In light of Lemma 4.14 this can be restated as saying that c_1 and c_2 are equivalent if and only if the geodesic rays $A(c_1)$ and $A(c_2)$ are equivalent.

With this in mind, we can think of $\partial\mathbb{H}^n$ as the set of equivalence classes of quasi-geodesic rays $c : [0, \infty) \to \mathbb{H}^n$. As such we denote points of $\partial\mathbb{H}^n$ by $[c]$. Having this extra flexibility allows us to extend quasi-isometries to maps from $\partial\mathbb{H}^n$ to itself in a straightforward way.

Notice that if we fix a basepoint $p \in \mathbb{H}^n$ there is a unique geodesic ray $\gamma : [0, \infty) \to \mathbb{H}^n$ with $\gamma(0) = p$ in each equivalence class of quasi-geodesic rays. Hence we can still define the topology on $\partial\mathbb{H}^n$ as we did in Section 2.1.2.

4.3.3 Quasi-isometries of $\mathbb{H}^n$ induce homeomorphisms of $\partial\mathbb{H}^n$.

Lemma 4.16. *If X and Y are metric spaces and $f : X \to Y$ is a (λ, ϵ)-quasi-isometry then if $A, B \subset X$, $d_{\mathcal{H}}(f(A), f(B)) \leq \lambda d_{\mathcal{H}}(A, B) + \epsilon$.*

Proof. Let $d_{\mathcal{H}}(A, B) = R$. Then for every $a \in A$ there is some $b_a \in B$ such that $d_X(a, b_a) \leq R$, and for every $b \in B$ there is some $a_b \in A$ such that $d_X(b, a_b) \leq R$. Since f is a quasi-isometry, for every $f(a) \in f(A)$ and every $f(b) \in f(B)$,

$$d_Y(f(a), f(b_a)) \leq \lambda d_X(a, b_a) + \epsilon \leq \lambda R + \epsilon \quad \text{and} \quad d_Y(f(b), f(a_b)) \leq \lambda d_X(b, a_b) + \epsilon \leq \lambda R + \epsilon,$$

establishing that $d_{\mathcal{H}}(f(A), f(B)) \leq \lambda R + \epsilon$. □

Proposition 4.17. *If $f : \mathbb{H}^n \to \mathbb{H}^n$ is a (λ, ϵ)-quasi-isometry then $\partial f : \partial\mathbb{H}^n \to \partial\mathbb{H}^n$, where*

$$\partial f([c]) = [f \circ c],$$

is a bijection with inverse given by ∂g where $g : \mathbb{H}^n \to \mathbb{H}^n$ is a quasi-inverse of f.

Proof. To see that ∂f is well defined, suppose $c_1, c_2 : [0, \infty) \to \mathbb{H}^n$ are quasi-geodesic rays with $[c_1] = [c_2]$. Then since f is a quasi-isometry and the Hausdorff distance between the images of c_1 and c_2 is finite, the hypotheses of Lemma 4.16 are satisfied. Hence the Hausdorff distance between the images of $f \circ c_1$ and $f \circ c_2$ is also finite and so $\partial f([c_1]) = \partial f([c_2])$.

Suppose $g : \mathbb{H}^n \to \mathbb{H}^n$ is a κ-quasi-inverse of $\mathbb{H}^n \to \mathbb{H}^n$. Then if $[c] \in \partial\mathbb{H}^n$, it is obvious that the Hausdorff distances between the image of c and the images of $g \circ f \circ c$ and $f \circ g \circ c$ are each at most κ. So it follows that

$$[c] = [f \circ g \circ c] = (\partial f \circ \partial g)([c]) \quad \text{and} \quad [c] = [g \circ f \circ c] = (\partial g \circ \partial f)([c]).$$

So ∂f and ∂g are mutually inverse. □

From now on, given any quasi-isometry $f : \mathbb{H}^n \to \mathbb{H}^n$ we will use the notation $\partial f : \partial\mathbb{H}^n \to \partial\mathbb{H}^n$ to denote the induced map defined in Proposition 4.17. We will also use the notation $\bar{f} : \overline{\mathbb{H}}^n \to \overline{\mathbb{H}}^n$ to denote the 'union' of f and ∂f.

Our final aim for this section is to show that ∂f is a homeomorphism. A similar (although not identical) result appears in Chapter 5 of Thurston's notes [Thu79]. We will essentially follow the presentation of this argument given in [BP92].

The next lemma, although rather technical, will help us establish the continuity of ∂f by helping us understand what the images of hyperplanes under quasi-isometries look like.

Lemma 4.18. *Suppose β is a geodesic line in $\mathbb{H}^n$, H is a hyperbolic hyperplane orthogonal to β, and $f : \mathbb{H}^n \to \mathbb{H}^n$ is a (λ, ϵ)-quasi-isometry. Let $\pi_{A(f \circ \beta)} : \mathbb{H}^n \to \mathbb{H}^n$ denote orthogonal projection onto the geodesic line $A(f \circ \beta)$. Then there exists some constant c depending only on λ, ϵ such that*

$$diam(\pi_{A(f \circ \beta)} f(H)) \leq c.$$

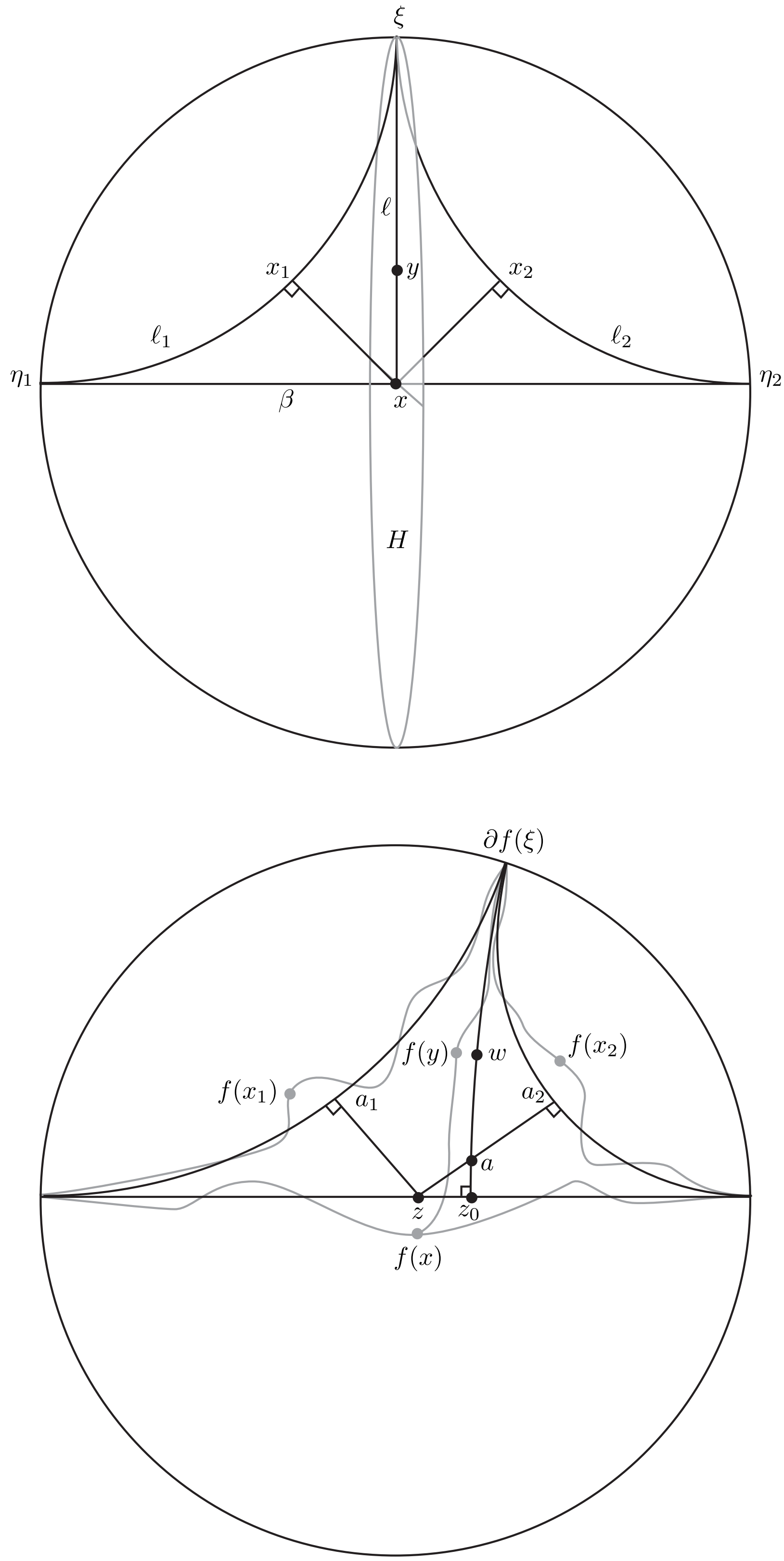


Figure 4.3: The top figure shows the construction used in the proof of Lemma 4.18. The bottom figure shows its image under $\bar{f}$.

Proof. The proof is based the geometric construction shown in Figure 4.3.

Let x be the point of intersection of β and H and take some $y \in H \setminus \{x\}$. Let ℓ be the geodesic ray contained in H that starts at x and passes through y. Let ξ be the endpoint of ℓ and η_1 and η_2 the endpoints of β. Let ℓ_1 and ℓ_2 be the geodesic lines joining ξ and η_1 and η_2 respectively and let x_i be the point on ℓ_i closest to x (for $i = 1, 2$).

For $i = 1, 2$, $d(x, x_i) = k = \cosh^{-1}(\sqrt{2})$ is a fixed constant. To see this, by symmetry the geodesic segment $[x, x_i]$ divides the triangle ξ, x, η_i into two congruent pieces. Hence the angle $\xi x x_i$ is $\pi/4$. Applying the cosine rule for triangles with one finite side and one right-angle (2.2) gives $\cosh d(x, x_i) = 1/\sin(\pi/4) = \sqrt{2}$.

Now consider the image of the diagram under $\bar{f}$. Let z be the point on $A(f \circ \beta)$ that is closest to $f(x)$. Furthermore, let z_0 be the foot of the perpendicular from $\partial f(\xi)$ to $A(f \circ \beta)$. First note that z is a uniformly bounded distance from each $A(f \circ \ell_i)$ since

$$d(A(f \circ \ell_i), z) \leq d(A(f \circ \ell_i), f(x_i)) + d(f(x_i), f(x)) + d(f(x), z) \leq R + \lambda k + \epsilon + R.$$

(Here we have used the definition of $A(\cdot)$ and the fact that f is a quasi-isometry.)

Now let a_1 and a_2 denote the points on $A(f \circ \ell_1)$ and $A(f \circ \ell_2)$ respectively that are closest to z. One of the geodesic segments $[z, a_i]$ intersects the geodesic ray emanating from z_0 with endpoint $\partial f(\xi)$. Without loss of generality assume it is $[z, a_2]$ and let the point of intersection be a.

Then azz_0 is a right angled hyperbolic triangle so

$$d(z, z_0) \leq d(z, a) \leq d(z, A(f \circ \ell_2)) \leq 2R + \lambda k + \epsilon.$$

Now suppose $w \in A(f \circ \ell)$. Then the projection of w onto $A(f \circ \beta)$ lies on the geodesic segment $[z, z_0]$. So $d(\pi_{A(f\circ\beta)}(w), z) \leq 2R + \lambda k + \epsilon$. If w is the closest point on $A(f \circ \ell)$ to $f(y)$ then, because orthogonal projection reduces distances,

$$\begin{aligned} d(\pi_{A(f\circ\beta)}(f(y)), z) &\leq d(\pi_{A(f\circ\beta)}(f(y)), \pi_{A(f\circ\beta)}(w)) + d(\pi_{A(f\circ\beta)}(w), z) \\ &\leq d(f(y), w) + d(\pi_{A(f\circ\beta)}(w), z) \\ &\leq R + 2R + \lambda k + \epsilon. \end{aligned}$$

So taking $c = 2(3R + \lambda k + \epsilon)$ completes the proof. □

Let us briefly recall the way we defined the topology on $\partial\mathbb{H}^n$. Given any geodesic ray β and some point y on β, let H be the hyperplane orthogonal to β and passing through y. Let Q_y denote the component of $\overline{\mathbb{H}}^n \setminus H$ containing the point $[\beta] \in \partial\mathbb{H}^n$. Then the Q_y are a basis of neighbourhoods at $[\beta]$.

Proposition 4.19. *If $f : \mathbb{H}^n \to \mathbb{H}^n$ is a quasi-isometry then $\partial f : \partial\mathbb{H}^n \to \partial\mathbb{H}^n$ is a homeomorphism.*

Proof. We have already seen that ∂f is a bijection whose inverse is ∂g where $g : \mathbb{H}^n \to \mathbb{H}^n$ is any quasi-inverse of f. Hence, by symmetry, we need only show that ∂f is continuous. We will do so by showing that $\bar{f} : \overline{\mathbb{H}}^n \to \overline{\mathbb{H}}^n$ is continuous at any $[\beta] \in \partial\mathbb{H}^n$.

As such, let β be a fixed geodesic ray in $\mathbb{H}^n$ with endpoint $[\beta] \in \partial\mathbb{H}^n$. Fix a neighbourhood Q of $\bar{f}([\beta])$ which is constructed using the geodesic ray $A(f \circ \beta)$. Then for sufficiently large $t_0 \in [0, \infty)$, for every $t > t_0$ the ball of radius c (where c is chosen as in the proof of Lemma 4.18) around $f(\beta(t))$ is contained in Q.

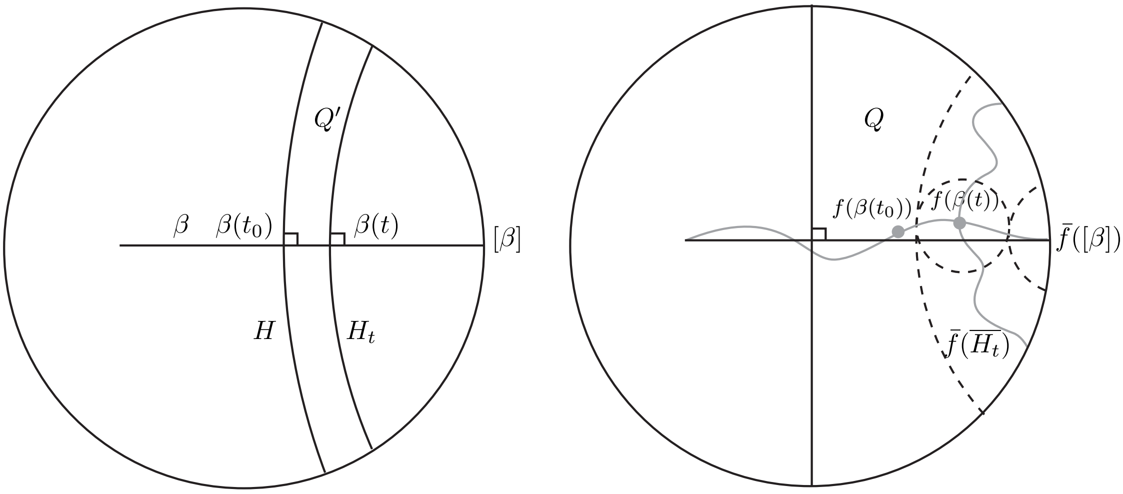


Figure 4.4: Construction showing that ∂f is continuous.

Let H be the hyperplane orthogonal to β passing through $\beta(t_0)$ and let Q' be the corresponding neighbourhood of $[\beta]$. For some $t > t_0$, let H_t be the hyperplane orthogonal to β passing through $\beta(t)$. Then applying Lemma 4.18 we have that $f(H_t) \subset Q$.

Taking the closure $\overline{H_t}$ of H_t, by the definition of $\bar{f}$ we also have that $\bar{f}(\overline{H_t}) \subset Q$. Since

$$Q' = \bigcup_{t \in [t_0, \infty)} \overline{H_t}$$

it follows that $\bar{f}(Q') \subset Q$ and so $\bar{f}$ is continuous at $[\beta]$. □

4.4 Proof of Theorem 4.1

Recall that our aim in this chapter is to prove the following.

Theorem 4.1. *Suppose Γ_1 and Γ_2 are subgroups of Isom$(\mathbb{H}^n)$ such that $M_1 = \mathbb{H}^n/\Gamma_1$ and $M_2 = \mathbb{H}^n/\Gamma_2$ are closed hyperbolic manifolds. If $\psi : \Gamma_1 \to \Gamma_2$ is an isomorphism then there is a homeomorphism $\partial f : \partial\mathbb{H}^n \to \partial\mathbb{H}^n$ such that*

$$\partial f \circ \gamma = \psi(\gamma) \circ \partial f \quad \textit{for all} \quad \gamma \in \Gamma_1.$$

The basic idea behind our proof of Theorem 4.1. is to construct a quasi-isometry from $\mathbb{H}^n$ to itself and use the definition of $\mathbb{H}^n$ in terms of quasi-geodesic rays to define a map on the boundary of hyperbolic space. To do this, we need a way to construct a quasi-isometry between $\mathbb{H}^n$ and a group of isometries that acts freely, properly discontinuously and cocompactly on $\mathbb{H}^n$ (and vice-versa).

The Švarc-Milnor Lemma provides us with a map in one direction. Taking any quasi-inverse of this map allows us to go back the other way.

Lemma 4.19 (Švarc-Milnor Lemma)**.** *Let X be a geodesic metric space. Suppose Γ acts properly discontinuously and co-compactly by isometries on X. Then Γ is finitely generated. If $\mathcal{A}$ is any* *finite generating set for* Γ *and* $x_0 \in X$ *is any basepoint, the map* $(\Gamma, d_{\mathcal{A}}) \ni \gamma \mapsto \gamma \cdot x_0 \in (X, d)$ *is a quasi-isometry.*

Again we defer the proof to Section 4.6. It is worth pointing out that this result actually holds for slightly more general spaces called 'length spaces' (for a definition see [BH99, p.32]).

Now we are ready to prove Theorem 4.1.

Proof of Theorem 4.1. Since $M_1 = \mathbb{H}^n/\Gamma_1$ and $M_2 = \mathbb{H}^n/\Gamma_2$ are compact manifolds it follows that Γ_1 and Γ_2 act properly discontinuously and co-compactly by isometries on $\mathbb{H}^n$.

Fix a basepoint $x_0 \in \mathbb{H}^n$. By the Švarc-Milnor Lemma, for $i = 1, 2$ the maps $\phi_i : \Gamma_i \to \mathbb{H}^n$ given by $\phi_i(\gamma) = \gamma \cdot x_0$ are quasi-isometries. Let $\rho_i : \mathbb{H}^n \to \Gamma_i$ denote fixed quasi-inverses of the ϕ_i defined as in the proof of Proposition 4.6. In particular, they satisfy $\phi_i \circ \rho_i \circ \phi_i = \phi_i$ for $i = 1, 2$. Since the ϕ_i are injective this implies that $\rho_i \circ \phi_i$ is the identity on Γ_i for $i = 1, 2$.

By Proposition 4.3 the isomorphism $\psi : \Gamma_1 \to \Gamma_2$ is a quasi-isometry. Since the composition of quasi-isometries is again a quasi-isometry it follows that $f : \mathbb{H}^n \to \mathbb{H}^n$ defined by

$$f = \phi_2 \circ \psi \circ \rho_1$$

is a quasi-isometry. So by Proposition 4.19 the map $\partial f : \partial\mathbb{H}^n \to \partial\mathbb{H}^n$ defined by $\partial f([c]) = [f \circ c]$ is a homeomorphism. It remains to show that ∂f satisfies the equivariance condition

$$(\partial f \circ \gamma)([c]) = \partial f([\gamma \circ c]) = [f \circ \gamma \circ c] = [\psi(\gamma) \circ f \circ c] = \psi(\gamma)([f \circ c]) = (\psi(\gamma) \circ \partial f([c]) \tag{4.2}$$

for all $\gamma \in \Gamma_1$ and all quasi-geodesic rays c. Of these, the only equality that requires proof is $[f \circ \gamma \circ c] = [\psi(\gamma) \circ f \circ c]$. So all that remains is to establish this relationship.

We first show that within every equivalence class of quasi-geodesic rays in $\mathbb{H}^n$, there is a quasi-geodesic of the form $c(t) = \eta(t) \cdot x_0$ where $\eta : [0, \infty) \to \Gamma_1$.

Take some quasi-geodesic ray $c' : [0, \infty) \to \mathbb{H}^n$. Since $\phi_1 \circ \rho_1 : \mathbb{H}^n \to \mathbb{H}^n$ is a quasi-isometry, $c = \phi_1 \circ \rho_1 \circ c'$ is a quasi-geodesic ray. Since ϕ_1 and ρ_1 are quasi-inverses of each other, $[c] = [c']$. Taking $\eta(t) = (\rho_1 \circ c')(t)$ it follows from the definition of ϕ_1 that $c(t) = \eta(t) \cdot x_0$, as required.

If $c(t) = \eta(t) \cdot x_0$ then $(\gamma \circ c)(t) = \gamma(\eta(t) \cdot x_0) = (\gamma\eta)(t) \cdot x_0$ (by the properties of group actions). Recall that $\rho_1 \circ \phi_1$ is the identity map on Γ_1. Hence

$$(\rho_1 \circ c)(t) = (\rho_1 \circ \phi_1 \circ \eta)(t) = \eta(t) \tag{4.3}$$

and, similarly,

$$(\rho_1 \circ \gamma \circ c)(t) = (\rho_1 \circ \phi_1 \circ \gamma\eta)(t) = \gamma\eta(t). \tag{4.4}$$

Then using the fact that ψ is a homomorphism,

$$\begin{aligned}
(f \circ \gamma \circ c)(t) &= (\phi_2 \circ \psi \circ \rho_1 \circ \gamma \circ c)(t) && \text{(by definition of } f\text{)} \\
&= \psi(\gamma\eta(t)) \cdot x_0 && \text{(by (4.4) and the definition of } \phi_2\text{)} \\
&= \psi(\gamma)\psi(\eta(t)) \cdot x_0 && \text{(since } \psi \text{ is a homomorphism)} \\
&= \psi(\gamma)(\psi(\eta(t)) \cdot x_0) && \text{(by properties of group actions)} \\
&= (\psi(\gamma) \circ (\phi_2 \circ \psi \circ \varphi_1 \circ c))(t) && \text{(by (4.3) and the definition of } \phi_2\text{)} \\
&= (\psi(\gamma) \circ f \circ c)(t) && \text{(by the definition of } f\text{).}
\end{aligned}$$

So the required equivariance property (4.2) holds, completing the proof. ∎

4.5 Quasi-isometric rigidity of lattices

Gromov's address at the 1983 International Congress of Mathemtatics proposed a bold and influential new program to study the properties of finitely generated groups by studying their geometric properties. A significant aspect of Gromov's program is to classify finitely generated groups up to quasi-isometry. One sub-program that has arisen from this is a problem of a similar nature to that solved by Mostow Rigidity-type theorems.

Suppose G is a Lie group. If Γ is a discrete subgroup of G such that the quotient G/Γ is compact, we call Γ a *co-compact lattice* in G. Quasi-isometric rigidity theorems typically state that for a particular Lie group G, if H is a subgroup of G that is quasi-isometric to a co-compact lattice of G, then H is (or is 'nearly') a co-compact lattice of G. In the case where $G = \mathrm{Isom}(\mathbb{H}^n)$ such a statement was proved by Cannon and Cooper in [CC92].

Theorem 4.20. *If H is a finitely generated group that is quasi-isometric to a co-compact lattice in Isom($\mathbb{H}^n$) (for $n \geq 3$) then H has a finite index subgroup that is itself a co-compact lattice in Isom($\mathbb{H}^n$).*

Results of this type have now been established for a number of Lie groups and for both co-compact lattices and lattices where the quotient is only required to have finite Haar measure (sometimes called non-uniform lattices). For much more on this problem see Farb's survey artice [Far97].

4.6 Proofs of the Švarc-Milnor and Morse lemmas

In this section we give proofs of the Švarc-Milnor Lemma and the Morse Lemma. Both proofs follow, fairly closely, the respective presentations in [BH99]. That is, Proposition 1.6, Theorem 1.7, and Lemma 1.11 of Chapter III.H for the Morse lemma and Lemmas 8.10, and 8.18 and Proposition 8.19 of Chapter I.8 for the Švarc-Milnor lemma. Having said that, some of our arguments are a little different, and occasionally we have supplied some extra details. Our motivation for giving these proofs is to give the reader a flavour of the style of arguments used in these coarse-geometric proofs.

Lemma 4.19 (Švarc-Milnor Lemma)**.** *Let X be a geodesic metric space. Suppose Γ acts properly discontinuously and co-compactly by isometries on X. Then Γ is finitely generated. If $\mathcal{A}$ is any finite generating set for Γ and $x_0 \in X$ is any basepoint, the map $(\Gamma, d_{\mathcal{A}}) \ni \gamma \mapsto \gamma \cdot x_0 \in (X, d)$ is a quasi-isometry.*

Proof. Since Γ acts co-compactly on X there is some compact (and hence bounded) $K \subset X$ such that $x_0 \in K$ and $X = \Gamma \cdot K := \bigcup_{\gamma\in\Gamma} \gamma \cdot K$. Let κ be the diameter of K. If $x \in X$ there is some $\gamma \in \Gamma$ such that $x \in \gamma \cdot K$. Hence $d_X(x, \gamma \cdot x_0) \leq \kappa$. Thus the map $\gamma \mapsto \gamma \cdot x_0$ is quasi-surjective.

Let $\mathcal{A} = \{\gamma \in \Gamma : d(x_0, \gamma \cdot x_0) \leq 4\kappa\}$. To see that $\mathcal{A}$ is finite, we argue by contradiction.

Suppose $\mathcal{A}$ is infinite. Then since the closed ball $B(x_0, 4\kappa)$ is compact in X and $\mathcal{A} \cdot \{x_0\}$ is a sequence in $B(x_0, 4\kappa)$, it follows by compactness that $\mathcal{A} \cdot \{x_0\}$ has an accumulation point. This means that the set $\{\gamma \in \mathcal{A} : B(x_0, \kappa) \cap \gamma \cdot B(x_0, \kappa) \neq \emptyset\}$ is infinite, contradicting the assumption that Γ acts properly discontinuously on X. So $\mathcal{A}$ is finite.

To see that $\mathcal{A}$ generates Γ, first let H be the subgroup of Γ generated by $\mathcal{A}$. Let U be an open neighbourhood of K of diameter at most 2κ. Then let $V = H \cdot U$ and $V' = (\Gamma \setminus H) \cdot U$ and note that $V \cup V' = X$. If $V \cap V' \neq \emptyset$ then there is some $x \in X$ and $\gamma \in H$, $\gamma' \notin H$ such that $d_X(\gamma \cdot x_0, x) \leq 2\kappa$ and $d_X(\gamma' \cdot x_0, x) \leq 2\kappa$. But then because γ acts by isometries on X,

$$d_X(x_0, \gamma^{-1}\gamma' \cdot x_0) = d_X(\gamma \cdot x_0, \gamma' \cdot x_0) \leq d_X(\gamma \cdot x_0, x) + d_X(x, \gamma' \cdot x_0) \leq 4\kappa.$$

So $\gamma^{-1}\gamma' \in H$ implying that $\gamma' \in H$, which is a contradiction. Hence $V' \cap V = \emptyset$. Since both V and V' are open, V is non-empty and X is connected, it follows that $V = X$. Hence $\mathcal{A}$ is a finite generating set for Γ.

Let $d_{\mathcal{A}}$ denote the word metric on Γ with respect to this generating set.

Let $\lambda = \max\{d(x_0, \gamma \cdot x_0) : \gamma \in \mathcal{A} \cup \mathcal{A}^{-1}\}$. Then if $\gamma, \gamma' \in \Gamma$ are such that $d_{\mathcal{A}}(\gamma, \gamma') = n$ we can write

$$\gamma^{-1}\gamma' = a_1 a_2 \cdots a_n \quad \text{where } a_i \in \mathcal{A} \cup \mathcal{A}^{-1}.$$

Let $\gamma_0 = 1$ and $\gamma_i = a_1 a_2 \cdots a_i$ for $i = 1, \ldots, n$. Then since Γ acts by isometries and $\gamma_{i-1}^{-1}\gamma_i = a_i$,

$$\begin{aligned} d_X(\gamma \cdot x_0, \gamma' \cdot x_0) &= d_X(x_0, \gamma^{-1}\gamma \cdot x_0) \\ &= d_X(x_0, a_1 a_2 \cdots a_n \cdot x_0) \\ &\leq \sum_{i=1}^{n} d_X(\gamma_{i-1} \cdot x_0, \gamma_i \cdot x_0) \\ &= \sum_{i=1}^{n} d_X(x_0, a_i \cdot x_0) \\ &\leq \lambda d_{\mathcal{A}}(\gamma, \gamma'). \end{aligned}$$

Now consider $\gamma \cdot x_0, \gamma' \cdot x_0 \in X$ and let $c : [0,1] \to X$ be the geodesic segment joining them. Let $d = d_X(\gamma \cdot x_0, \gamma' \cdot x_0)$ and let $N = \lceil d/\kappa \rceil$. Partition $c([0,1])$ into geodesic segments of length at most κ by choosing $0 = t_0 < t_1 < \cdots < t_{N-1} < t_N = 1$ such that $d(c(t_{i-1}), c(t_i)) = \kappa$ for $i = 1, \ldots N-1$. Note that our choice of N ensures that $d(c(t_{N-1}), c(t_N))) \leq \kappa$.

Recall that $\gamma \mapsto \gamma \cdot x_0$ is quasi-surjective. Hence for $i = 0, \ldots, N$ there is some $\gamma_i \in \Gamma$ such that $d_X(\gamma_i \cdot x_0, c(t_i)) \leq \kappa$. (Furthermore we can choose $\gamma_0 = \gamma$ and $\gamma_N = \gamma'$). Since

$$d_X(\gamma_i \cdot x_0, \gamma_{i+1} \cdot x_0) \leq d_X(\gamma_i \cdot x_0, c(t_i)) + d_X(c(t_i), c(t_{i+1})) + d_X(c(t_{i+1}), \gamma_{i+1} \cdot x_0) \leq 3\kappa$$

it follows that $\gamma_i^{-1}\gamma_{i+1} \in \mathcal{A}$. Since we can write

$$\gamma^{-1}\gamma' = (\gamma_0^{-1}\gamma_1)(\gamma_1^{-1}\gamma_2) \cdots (\gamma_{N-1}^{-1}\gamma_N)$$

it follows from the definition of N that

$$d_{\mathcal{A}}(\gamma, \gamma') \leq N \leq \frac{1}{\kappa} d_X(\gamma \cdot x_0, \gamma' \cdot x_0).$$

So $\gamma \mapsto \gamma \cdot x_0$ is also a quasi-isometric embedding. $\square$

4.6.1 Proof of the Morse lemma

In our proof of the Morse Lemma, we will make the following standard definition of the length of a path in a metric space.

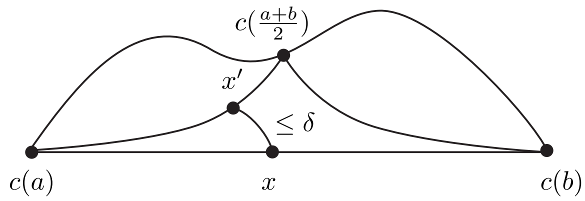


Figure 4.5: The induction step for lemma 4.22.

Definition 4.21. If X is a metric space then a path $c : [a, b] \to X$ is called *rectifiable* if

$$\ell(c) = \sup \sum_i d(c(a_i, a_{i+1}))$$

is finite (where the supremum is taken over all finite subdivisions $a = a_0 < a_1 < \cdots < a_N = b$ of $[a, b]$). If a path c is rectifiable then we define its *length* to be $\ell(c)$.

In a geodesic metric space, the distance between any two points is precisely the length of the geodesic segment joining them. Before proving the Morse lemma we need two preliminary results.

Lemma 4.22. *Suppose X is a δ-hyperbolic space and $c : [a, b] \to X$ is a path in X. Then for any $x \in [c(a), c(b)]$,*

$$d(x, c([a, b])) \leq \delta|\log_2(\ell(c))| + 1.$$

Proof. Throughout, assume that c is parametrized by arc length.

If $\ell(c) \leq 2^0 = 1$ then $d(c(a), c(b)) \leq 1$ so

$$d(x, c([a, b])) \leq d(x, c(a)) \leq 1 \leq \delta|\log_2(\ell(c))| + 1.$$

Arguing by induction, suppose that the statement is true for any path c' with $\ell(c') \leq 2^N$ and assume that $\ell(c) \leq 2^{N+1}$. Then consider the geodesic triangle with vertices at $c(a), c(b)$ and $c((a + b)/2)$. (See Figure 4.5.) Then there is some x' either on the geodesic segment $[c(a), c((a+b)/2)]$ or the geodesic segment $[c((a+b)/2), c(b)]$ such that $d(x, x') \leq \delta$. Without loss of generality assume we are in the first situation and let c' be the restriction of c to $[a, (a+b)/2]$. Then $\ell(c') = \ell(c)/2 \leq 2^N$ and $x' \in [a, (a + b)/2]$. So by the induction hypothesis

$$d\left(x', c'([a, (a + b)/2])\right) \leq \delta|\log_2(\ell(c'))| + 1.$$

Then

$$\begin{aligned} d(x, c([a, b])) \leq d(x, x') + d(x', c([a, b])) &\leq d(x, x') + d(x', c'([a, (a + b)/2])) \\ &\leq (\delta + \delta|\log_2(\ell(c'))|) + 1 \\ &= \delta|\log_2(2\ell(c'))| + 1 = \delta|\log_2(\ell(c))| + 1 \end{aligned}$$

and we are done. □

The second result essentially says that for the purposes of proving the Morse lemma, we can replace quasi-geodesics with continuous quasi-geodesics and avoid having to worry about a whole host of pathological behaviour.

Since the details of the proof of this result are not particularly enlightening, we will just give the main construction but do not prove that it works.

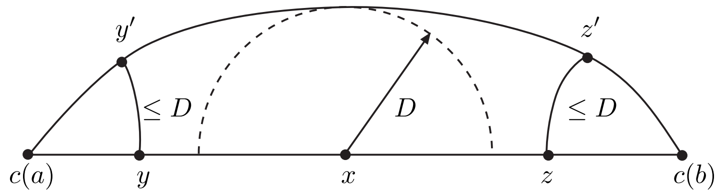


Figure 4.6: Images of geodesics under pseudo-isometries are close to geodesics.

Lemma 4.23 (Taming quasi-geodesics)**.** *If X is a geodesic metric space and $c : [a, b] \to X$ is a (λ, ϵ)-quasi-geodesic, then there are constants k_1, k_2, r and ϵ' (depending only on λ and ϵ) and there is a continuous (λ, ϵ')-quasi-geodesic $c' : [a, b] \to X$ such that*

1. *$c'(a) = c(a)$ and $c'(b) = c(b)$;*
2. *$\ell(c'|_{[t,t']}) \leq k_1 d(c'(t), c'(t')) + k_2$ for all $t, t' \in [a, b]$;*
3. *the Hausdorff distance between the images of c and c' is at most r.*

Sketch of proof. Let $N = \lceil a \rceil$ and $M = \lfloor b \rfloor$. Then let c' be given by concatenating the geodesic segments $[c(a), c(N)], [c(N), c(N+1)], [c(N+1), c(N+2)], \ldots, [c(M-1), c(M)], [c(M), c(b)]$ and linearly reparametrizing each segment appropriately.

It turns out that c' has all the desired properties. See [BH99] for the details. □

Lemma 4.13 (Morse Lemma)**.** *If X is a δ-hyperbolic space then there is some constant $R = R(\delta, \lambda, \epsilon)$ such that for any (λ, ϵ)-quasi-geodesic $\bar{c} : [a, b] \to X$,*

$$d_{\mathcal{H}}(\bar{c}([a, b]), [\bar{c}(a), \bar{c}(b)]) \leq R.$$

Proof. First, we replace $\bar{c}$ with a (λ, ϵ')-quasi-geodesic c 'tamed' by the method given in the sketched proof of Lemma 4.23.

Since $[c(a), c(b)]$ is compact there is some $x \in [c(a), c(b)]$ such that

$$d(x, c([a, b])) = \max_{y \in [c(a), c(b)]} d(y, c([a, b])).$$

Define D to be this maximum distance. Choose $y, z \in [c(a), c(b)]$ such that

$$D \leq d(y, c([a, b])) \leq 2D \quad \text{and} \quad D \leq d(z, c([a, b])) \leq 2D.$$

Let $y' = c(a'), z' = c(b')$ be points on $c([a, b])$ that are closest to y and z respectively and note that $d(y, y') \leq D$ and $d(z, z') \leq D$ by the definition of D. (See Figure 4.6.)

Let γ be the path from y to z that traverses the geodesic segment $[y, y']$ then follows c from y' to z' then traverses the geodesic segment $[z', z]$. Then, applying Lemma 4.23 we have that

$$\begin{aligned}\ell(\gamma) = d(y, y') + d(z, z') + \ell(c|_{[a',b']}) &\leq D + D + k_1 d(y', z') + k_2 \\ &\leq 2D + k_1(d(y, y') + d(y, z) + d(z, z')) + k_2 \\ &\leq 2D + 6k_1 D + k_2.\end{aligned}$$

Hence by Lemma 4.22

$$D = d(x, c([a, b])) \leq \delta |\log_2(2D + 6k_1 D + k_2)| + 1.$$

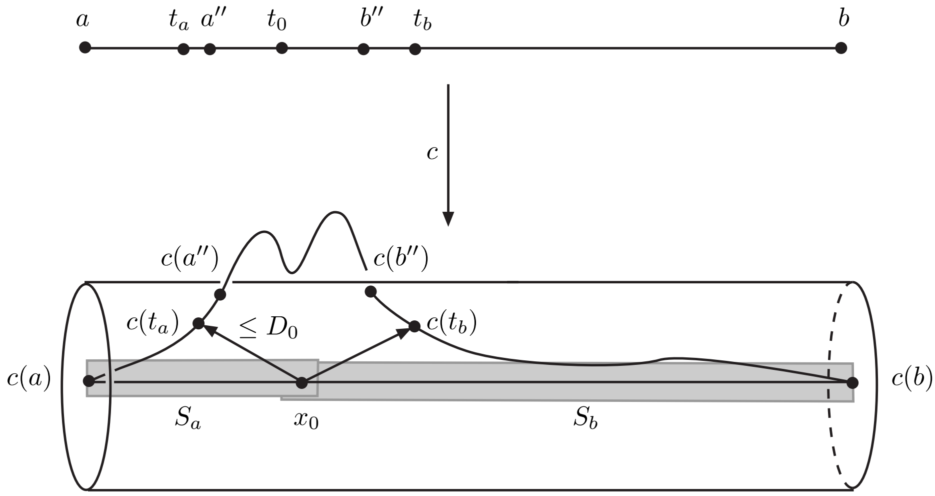


Figure 4.7: Images of geodesics under pseudo-isometries are close to geodesics.

Since k_1 and k_2 depend only on λ and ϵ this gives rise to an upper bound on D in terms of λ, ϵ and δ which we call D_0. Observe that $[c(a), c(b)] \subseteq \mathcal{N}_{D_0}(c([a, b]))$.

For the remainder of the proof, refer to Figure 4.7.

To show that the Hausdorff distance is uniformly bounded it remains to find some $R = R(\delta, \lambda, \epsilon)$ such that $c([a, b]) \subseteq \mathcal{N}_R([c(a), c(b)])$. If $c([a, b]) \subseteq \mathcal{N}_{D_0}([c(a), c(b)])$ the choosing $R = D_0$ completes the proof. So, assume this is not the case.

Let $[a'', b'']$ be maximal such that $c([a'', b''])$ is contained in the complement of the open D_0 neighbourhood of $[c(a), c(b)]$. Let

$$S_a = \{x \in [c(a), c(b)] : d(x, c([a, a''])) \leq D_0\} \quad \text{and} \quad S_b = \{x \in [c(a), c(b)] : d(x, c([b'', b])) \leq D_0\}.$$

These are clearly closed and non-empty and cover the connected set $[c(a), c(b)]$. Hence they are non-disjoint so we can choose $x_0 \in S_a \cap S_b$, $t_a \in [a, a'']$ and $t_b \in [b'', b]$ such that $d(x_0, c(t_a)) \leq D_0$ and $d(x_0, c(t_b)) \leq D_0$. Then by the triangle inequality $d(c(t_a), c(t_b)) \leq 2D_0$, and so since c is a (λ, ϵ')-quasi-isometric embedding, $d(t_a, t_b) \leq 2\lambda D_0 + \lambda\epsilon'$.

Now take any $t_0 \in [a'', b'']$. Then

$$\begin{aligned} d(c(t_0), [c(a), c(b)]) &\leq d(c(t_0), c(t_a)) + d(c(t_a), [c(a), c(b)]) \\ &\leq \lambda d(t_0, t_a) + \epsilon' + D_0 \\ &\leq \lambda d(t_a, t_b) + \epsilon' + D_0 \\ &\leq 2\lambda^2 D_0 + \lambda^2\epsilon' + \epsilon' + D_0 =: R' \end{aligned}$$

Since ϵ' depends only on λ and ϵ and D_0 depends only on λ, ϵ, and δ, it follows that R' depends only on λ, ϵ, and δ.

Since c and $\bar{c}$ have the same endpoints, $[\bar{c}(a), \bar{c}(b)] = [c(a), c(b)]$. Furthermore, since the Hausdorff distance between $\bar{c}([a, b])$ and $c([a, b])$ is at most r, it follows that

$$d_{\mathcal{H}}([\bar{c}(a), \bar{c}(b)], \bar{c}([a, b])) \leq d_{\mathcal{H}}([c(a), c(b)], c([a, b])) + d_{\mathcal{H}}(c([a, b]), \bar{c}([a, b])) \leq R' + r.$$

Because r only depends on λ and ϵ, choosing $R = R' + r$ completes the proof. $\square$

Chapter 5

The proof of Besson, Courtois, and Gallot

In this chapter we outline the proof of Mostow Rigidity given by Besson, Courtois, and Gallot in their 1996 paper [BCG96]. The most striking feature of their proof is that it is constructive — given two closed hyperbolic manifolds with isomorphic fundamental groups, Besson et al. construct an isometry between them!

From Chapter 4 we know that an isomorphism between fundamental groups of closed hyperbolic n-manifolds gives rise to a homeomorphism ∂f of $\partial\mathbb{H}^n$. Building on the work of Douady and Earle [DE86], Besson et al. describe a method of extending ∂f to a smooth map F of $\mathbb{H}^n$ which is often called the 'barycentric extension' of ∂f.

Remarkably, it is possible to give tight estimates on the Jacobian of F. In particular it turns out that the barycentric extension is always volume non-increasing in the sense that the Jacobian at every point is at most one. If the barycentric extension map, F, is a map of degree one it can be shown that the Jacobian of F is exactly one at each point. Then a careful analysis of this equality case shows that F is actually an isometry.

In Section 5.1 we introduce Busemann functions, which give a way to measure the 'distance' between points in $\mathbb{H}^n$ points in $\partial\mathbb{H}^n$. This will allow us to introduce 'visual measures' and the 'barycentre of a measure' — two notions that play a major role in the construction of the barycentric extension map. The former gives a nice way to associate with each $x \in \mathbb{H}^n$ a probability measure μ_x on $\partial\mathbb{H}^n$. The latter gives a way to associate with (almost) any probability measure ν on $\partial\mathbb{H}^n$ a point $\mathrm{bar}\,(\nu) \in \mathbb{H}^n$. These constructions will be the subject of Sections 5.2 and 5.3.

In Section 5.4 we describe the construction of the barycentric extension map and its most important properties. Finally in Section 5.5 we combine the results of Chapter 4 and this chapter to give a complete proof of Mostow's Strong Rigidity Theorem for closed hyperbolic manifolds.

The proof of Mostow Rigidity given by Besson, Courtois, and Gallot is actually a special case of a more general result. In Section 5.6 we outline this more general viewpoint, and show how Mostow Rigidity arises as a special case.

5.1 Busemann functions

It makes no sense to define the 'distance' between a point in $\mathbb{H}^n$ and a point in $\partial\mathbb{H}^n$. But we can define a sensible notion of 'relative distance' between points in $\mathbb{H}^n$ and a point in $\partial\mathbb{H}^n$. Busemann functions capture this notion nicely.

Definition 5.1. Given a point $O \in \mathbb{H}^n$, the *Busemann function normalized at O* is the function $B_O : \mathbb{H}^n \times \partial\mathbb{H}^n \to \mathbb{R}$ given by

$$B_O(x, [\beta]) := \lim_{t\to\infty} d(\beta(t), x) - d(\beta(t), O)$$

where the representative β is chosen to satisfy $\beta(0) = O$.

This definition holds for much more general spaces than hyperbolic spaces. In $\mathbb{H}^n$ there is an equivalent definition that is more geometric in nature.

Proposition 5.2. *Suppose $O, x \in \mathbb{H}^n$ and $\xi \in \partial\mathbb{H}^n$ and let H denote the horosphere passing through ξ and O. Then*

$$B_O(x, \xi) = \pm d(x, H)$$

where the sign is chosen according to whether x is inside (minus sign) or outside (plus sign) the horoball whose boundary is H.

We will not prove this, but at this stage a remark is in order about the definition of 'horosphere'. In Section 2.1.1 we gave a geometric definition of horosphere that is quite appealing in $\mathbb{H}^n$. A more intrinsic (and general) definition is to say that horospheres are the level sets of Busemann functions. From this point of view the previous proposition is a tautology.

From now on we will assume that our Busemann functions are always normalized at the origin in the Poincaré ball model and at the equivalent point in the other models. Throughout we will denote this point by $O \in \mathbb{H}^n$, but will omit it from our notation for the Busemann function, writing $B : \mathbb{H}^n \times \partial\mathbb{H}^n \to \mathbb{R}$ instead. Furthermore, we will write either $B_x : \partial\mathbb{H}^n \to \mathbb{R}$ or $B_\xi : \mathbb{H}^n \to \mathbb{R}$ whenever we want to think of B as a function of one variable with the other left fixed.

Proposition 5.3.

1. *If $\varphi \in Isom(\mathbb{H}^n)$ is a reflection in a hyperplane containing ξ then*

$$B(\varphi(x), \xi) = B(x, \xi) \quad \textit{and} \quad d_x B_\xi(\cdot) = (d_{\varphi(x)} B_\xi \circ d_x\varphi)(\cdot).$$

2. *If $\varphi \in Isom(\mathbb{H}^n)$ fixes O then*

$$B(\varphi(x), \varphi(\xi)) = B(x, \xi) \quad \textit{and} \quad d_x B_\xi(\cdot) = (d_{\varphi(x)} B_{\varphi(\xi)} \circ d_x\varphi)(\cdot).$$

3. *$B_\xi : \mathbb{H}^n \to \mathbb{R}$ is smooth for any $\xi \in \partial\mathbb{H}^n$;*

4. *$\nabla B_\xi(x)$ is the unit vector in $T_x\mathbb{H}^n$ parallel to the geodesic line joining x and ξ and pointing away from ξ.*

5. *The Hessian of B_ξ is given by:*

$$Hess_x(B_\xi)(u, v) = \langle u, v\rangle - \langle \nabla B_\xi(x), u\rangle \langle \nabla B_\xi(x), v\rangle .$$

6. *For all $u \in T_x\mathbb{H}^n$, $Hess_x(B_\xi)(u, u) \geq 0$ with equality if and only if u is parallel to $\nabla B_\xi(x)$.*

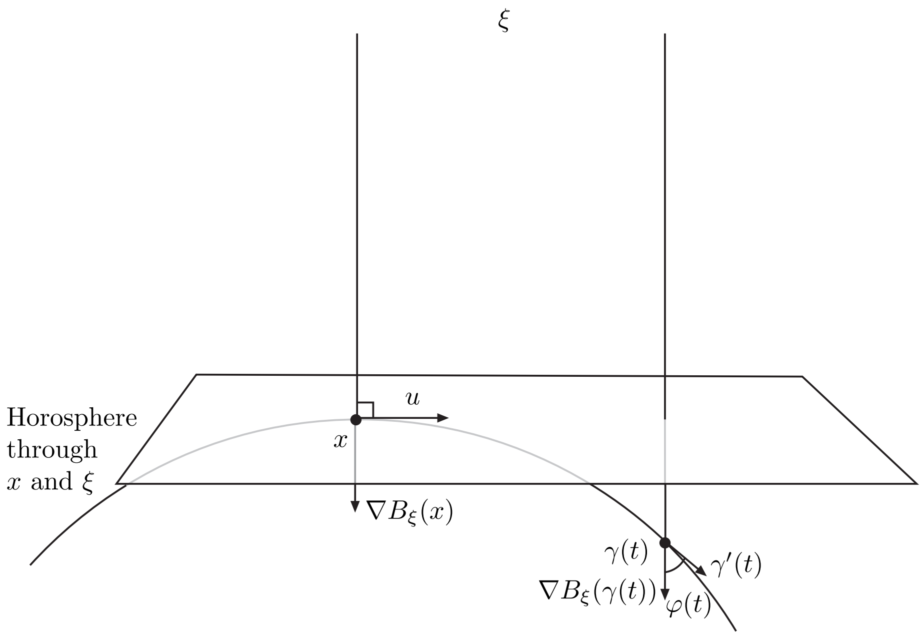


Figure 5.1: Computing the Hessian of the Busemann function.

Proof.

1. Let H denote the horosphere passing through ξ and O. Any reflection φ in a hyperplane containing ξ must satisfy $\varphi(H) = H$ for all horospheres H passing through ξ. Then if x is inside (outside) the horoball whose boundary passes through ξ and O then $\varphi(x)$ is also inside (outside) this horoball. Hence

$$B(\varphi(x), \xi) = \pm d(H, \varphi(x)) = \pm d(\varphi(H), \varphi(x)) = \pm d(H, x) = B(x, \xi).$$

The second assertion follows by the chain rule.

2. Let $\beta : [0, \infty) \to \mathbb{H}^n$ be a geodesic ray such that $\beta(0) = O$. Then since $\varphi(O) = O$ clearly $(\varphi \circ \beta)(0) = O$ so

$$\begin{aligned} B_O(\varphi(x), \varphi([\beta])) &= \lim_{t\to\infty} d(\varphi(\beta(t)), \varphi(x)) - d((\varphi \circ \beta)(t), O) \\ &= \lim_{t\to\infty} d(\beta(t), x) - d(\beta(t), \varphi^{-1}(O)) = B_O(x, [\beta]). \end{aligned}$$

The second assertion follows by the chain rule.

3. Consider the upper half-space model. Note that in this model $O = (0, 0, \ldots, 1)$. Using part 2 above, by composing with a 'rotation' about O, assume $\xi = \infty$. Then the horosphere through x is the Euclidean plane $H = \{y \in \mathbb{R}^n : y_n = 1\}$. If $x = (x_1, x_2, \ldots, x_n) \in \mathcal{H}^n$ then the distance from x to H is $|\ln(x_n)|$. To adapt to the sign convention chosen for B_ξ it follows that $B_\infty(x) = -\ln(x_n)$, which is certainly smooth.

4. Again let us work in the half-space model. Let $\gamma(t)$ be a smooth curve with $\gamma(0) = x \in \mathcal{H}^n$ and $\gamma'(0) = v \in T_x\mathbb{H}^n$. Then since $B_\infty(x) = -\ln(x_n)$ we have that

$$\langle \nabla B_\infty(x), v \rangle = \frac{d}{dt}\, B_\infty(\gamma(t))\big|_{t=0} = -\frac{v_n}{x_n} = \langle -x_n e_n, v \rangle\,.$$

So $\nabla B_\infty = -x_n e_n$ is the unit vector parallel to the geodesic joining ∞ and x and pointing away from ∞.

For arbitrary $\xi \in \partial\mathbb{H}^n$, $\nabla B_\xi(x)$ is just the image of $-x_n e_n$ under the (orientation preserving) isometry that fixes O and sends ∞ to ξ. Since our description of ∇B_ξ is invariant under such isometries, we are done.

5. Recall that since $\mathrm{Hess}_x(B_\xi)(\cdot,\cdot)$ is a symmetric bilinear form it is completely determined by the associated quadratic form.

   To compute $\mathrm{Hess}_x(B_\xi)(u,u)$ let γ be a geodesic with $\gamma(0) = x$ and $\gamma'(0) = u$. Then if $\varphi(t)$ is the angle between $\gamma'(t)$ and $\nabla B_\xi(\gamma(t))$, by (2.6), one of our methods for computing the Hessian, we have that

$$\mathrm{Hess}_x(B_\xi)(u,u) = \frac{d}{dt}\left\langle \nabla_{\gamma(t)} B_\xi, \gamma'(t)\right\rangle\bigg|_{t=0} = \frac{d}{dt}\cos\varphi(t)\bigg|_{t=0} = -\varphi'(0)\sin\varphi(0).$$

   If u is parallel to $\nabla_x B_\xi$ then $\varphi(0) = 0$ so the Hessian is zero.

   If u is orthogonal to $\nabla_x B_\xi$ then $\varphi(0) = \pi/2$ and by the hyperbolic cosine rule (2.1), at time t we have

$$\sin\varphi(t) = \mathrm{sech}(t).$$

   So

$$[\varphi'(t)\cos\varphi(t)]^2 = \left[\frac{d}{dt}\left(\sin\varphi(t)\right)\right]^2 = \left[\frac{d}{dt}\mathrm{sech}(t)\right]^2 = [-\mathrm{sech}(t)\tanh(t)]^2.$$

   Now $\cos^2\varphi(t) = 1 - \sin^2\varphi(t) = \tanh^2(t)$ so $\varphi'(0)^2 = \mathrm{sech}^2(0) = 1$. It remains to check whether $\varphi(t)$ is increasing or decreasing at $t = 0$. Since the angle sum of a hyperbolic triangle is at most π it follows that $\varphi(t) \leq \pi/2$ for all t. Furthermore, $\varphi(0) = \pi/2$ so $\varphi(t)$ is decreasing. Thus we conclude that $\varphi'(0) = -1$ and so $\mathrm{Hess}_x(B_\xi)(u,u) = 1$.

   From these two cases it is clear that

$$\mathrm{Hess}_x(B_\xi)(u,u) = \langle u,u\rangle - \langle \nabla B_\xi(x), u\rangle \langle \nabla B_\xi(x), u\rangle$$

   on a basis of $T_x\mathbb{H}^n$ and so this holds for all $u \in T_x\mathbb{H}^n$.

6. Simply observe that if $u \in T_x\mathbb{H}^n$,

$$\mathrm{Hess}_x(B_\xi)(u,u) = \langle u,u\rangle - \langle \nabla B_\xi(x), u\rangle^2 \geq \langle u,u\rangle - \langle u,u\rangle \langle \nabla B_\xi(x), \nabla B_\xi(x)\rangle$$

   with equality if and only if u and $\nabla B_\xi(x)$ are parallel (by the Cauchy-Schwarz inequality). Since $\nabla B_\xi(x)$ is a unit vector for all $x \in \mathbb{H}^n$ and $\xi \in \partial\mathbb{H}^n$ the result follows. □

5.2 The visual map and visual measures

5.2.1 The visual map

Recall that we can think of $\partial\mathbb{H}^n$ as being identified with $UT_O\mathbb{H}^n$. We will think of the unit tangent space to $\mathbb{H}^n$ at x as the 'visual sphere' — the set of directions in which we can look when standing at x. The visual map $V_x : UT_O\mathbb{H}^n \to UT_x\mathbb{H}^n$ keeps track of how what we see differs from what someone standing at O would see. That is, if a person standing at O shot a beam of light out in a particular direction u, if we are standing at x it will appear to converge to a point $V_x(u)$ in our visual sphere.

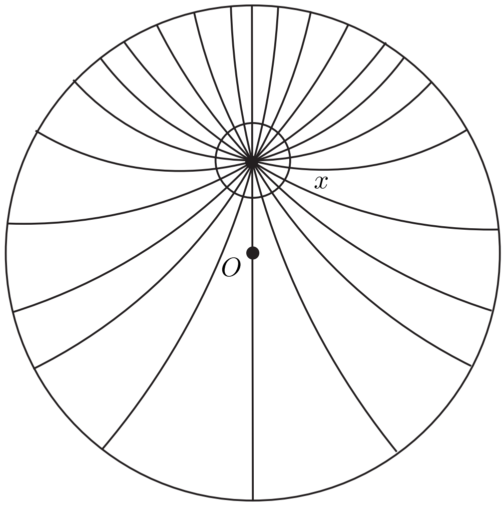


Figure 5.2: The visual map. The small sphere around x is the unit sphere in $T_x\mathbb{H}^n$. The outer sphere could be thought of as the boundary of the Poincaré model or the unit sphere in $T_O\mathbb{H}^n$. Note the variation in 'density' of the endpoints of the geodesic rays.

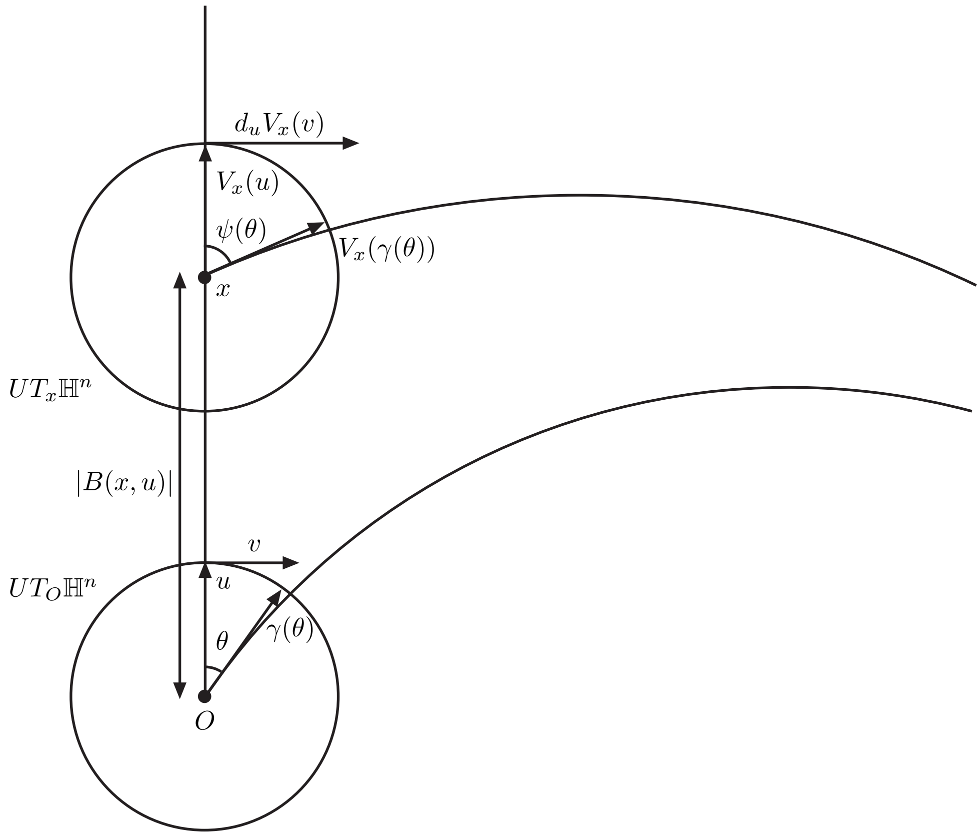


Figure 5.3: Computing the derivative of the visual map (in the upper half-space model).

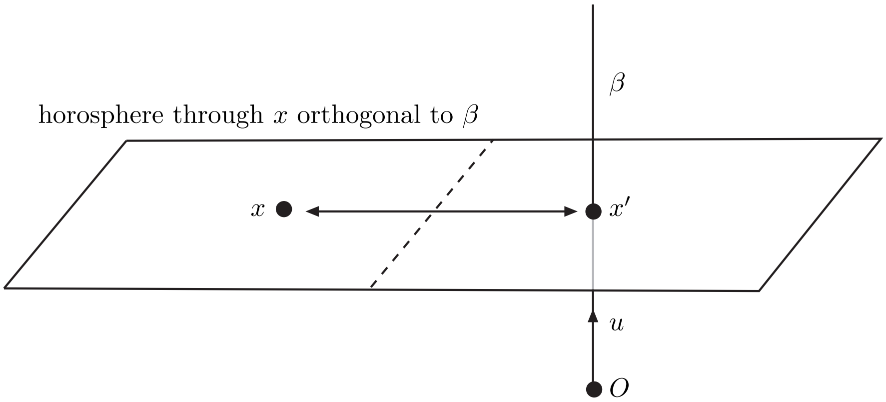


Figure 5.4: An isometry mapping x to x' where x' is on the geodesic ray β emanating from O with velocity u.

Given our description of the gradient of the Busemann function in Proposition 5.3, it follows that if $u \in UT_O\mathbb{H}^n$ then

$$V_x(u) = -\nabla B_u(x)$$

where by writing B_u we are silently making the identification between $UT_O\mathbb{H}^n$ and $\partial\mathbb{H}^n$.

Proposition 5.4. *If $V_x : UT_O\mathbb{H}^n \to UT_x\mathbb{H}^n$ is the visual map then the absolute value of its Jacobian at $u \in UT_O\mathbb{H}^n$ is $|Jac_u(V_x)| = e^{(n-1)B(x,u)}$.*

Proof. Suppose x lies on the geodesic ray $\beta : [0, \infty) \to \mathbb{H}^n$ emanating from O with initial velocity u. Suppose $v \in T_uUT_O\mathbb{H}^n$ is a unit tangent vector to $UT_O\mathbb{H}^n$. Then the hyperbolic plane P determined by u and v contains x. Furthermore for some $\epsilon > 0$, the curve $\gamma : (-\epsilon, \epsilon) \to UT_O\mathbb{H}^n$ given by $\gamma(\theta) = u\cos(\theta) + v\sin(\theta)$ is such that the plane determined by u and $\gamma(\theta)$ is also P. Hence, by the definition of the visual map, the plane determined by $V_x(u)$ and $V_x(\gamma(\theta))$ is also P.

This immediately tells us that V_x is conformal. It remains to find the factor that V_x scales tangent vectors by, that is $\|d_uV_x(v)\|$.

Note that since $V_x \circ \gamma$ is restricted to lie in a circle in $T_{V_x(u)}UT_x\mathbb{H}^n$, it can be parametrized by

$$(V_x \circ \gamma)(\theta) = \frac{V_x(u)}{\|V_x(u)\|}\cos(\psi(\theta)) + \frac{d_uV_x(v)}{\|d_uV_x(v)\|}\sin(\psi(\theta))$$

where $\psi(\theta)$ is the angle between $V_x(u)$ and $V_x(\gamma(\theta))$ (see Figure 5.3). Hence

$$d_uV_x(v) = \frac{d}{d\theta}(V_x \circ \gamma)(\theta)\Big|_{\theta=0} = \frac{d_uV_x(v)}{\|d_uV_x(v)\|}\psi'(0)$$

so it remains to find $\psi'(0) = \|d_uV_x(v)\|$.

From Figure 5.3 observe that the geodesic rays determined by $u \in UT_O\mathbb{H}^n$, $\gamma(\theta) \in UT_O\mathbb{H}^n$ and $V_x(\gamma(\theta)) \in UT_x\mathbb{H}^n$ form an ideal triangle with one side of length $|B(x,u)|$. Hence by the hyerbolic cosine rule (2.2)

$$\cosh(|B(x,u)|) = \frac{1 - \cos(\theta)\cos(\psi(\theta))}{\sin(\theta)\sin(\psi(\theta))}.$$

So taking the limit as $\theta \to 0$ (by two applications of L'Hôpital's rule), we obtain

$$\cosh(|B(x,u)|) = \frac{1}{2}\left(\psi'(0) + \frac{1}{\psi'(0)}\right).$$

Taking care that our sign convention is correct we conclude that

$$\|d_u V_x(v)\| = e^{-B(x,u)}.$$

If $v_1, v_2, \ldots, v_{n-1}$ is an orthonormal basis for $T_u UT_O\mathbb{H}^n$ then $V_x(v_1), \ldots, V_x(v_{n-1})$ is an orthonormal basis for $T_{V_x(u)}UT_x\mathbb{H}^n$. Furthermore, the derivative of V_x is represented in these bases by

$$[d_u V_x] = e^{-B(x,u)} I_{n-1}.$$

So the Jacobian of V_x is just the determinant of this matrix, namely $e^{-(n-1)B(x,u)}$.

In the case where $x \in \mathbb{H}^n$ does not lie on the geodesic ray $\beta : [0,\infty) \to \mathbb{H}^n$ emanating from O with intial velocity u, then there is a reflection φ in a hyperplane containing $\beta(\infty)$ so that $\varphi(x)$ does lie on β (see Figure 5.4).

Then by Part 1 of Proposition 5.3 it follows that

$$V_x(u) = -\nabla B_u(x) = -d_{\varphi(x)}\varphi^{-1}(\nabla B_u(\varphi(x))) = d_{\varphi(x)}\varphi^{-1} \circ V_{\varphi(x)}(u).$$

Since composition with an isometry does not affect the absolute value of the Jacobian the result holds in this case also. □

5.2.2 Visual measures

Let $\mathcal{P}(\partial\mathbb{H}^n)$ be the space of probability measures on $\partial\mathbb{H}^n$. Since we will identify $\partial\mathbb{H}^n$ with $UT_O\mathbb{H}^n$, giving us a concrete representation of $\partial\mathbb{H}^n$ as a Euclidean unit sphere, we will actually think of $\mathcal{P}(\partial\mathbb{H}^n)$ as $\mathcal{P}(UT_O\mathbb{H}^n)$.

Then with each $x \in \mathbb{H}^n$ we can associate a measure $\mu_x \in \mathcal{P}(\partial\mathbb{H}^n)$ by using the visual map in the following way.

Definition 5.5. Suppose $x \in \mathbb{H}^n$ and λ_x is the canonical measure on the unit sphere $UT_x\mathbb{H}^n$. Then the *visual measure* at $x \in \mathbb{H}^n$ is

$$\mu_x = (V_x^{-1})_*[\lambda_x],$$

the pushforward of λ_x to $UT_O\mathbb{H}^n$ by V_x^{-1}.

The following summarizes the properties of the visual measure that we will use.

Proposition 5.6. *Suppose $x \in \mathbb{H}^n$.*

1. *If $\varphi \in Isom(\mathbb{H}^n)$ then $\mu_{\varphi(x)} = \varphi_*[\mu_x]$. Hence μ_x is invariant under isometries that fix x.*
2. *The visual measure μ_x is absolutely continuous with respect to μ_O and so has no atoms.*
3. *The Radon-Nikodym derivative of μ_x with respect to μ_O is*

$$\frac{d\mu_x}{d\mu_O}(\theta) = e^{-(n-1)B(x,\theta)}.$$

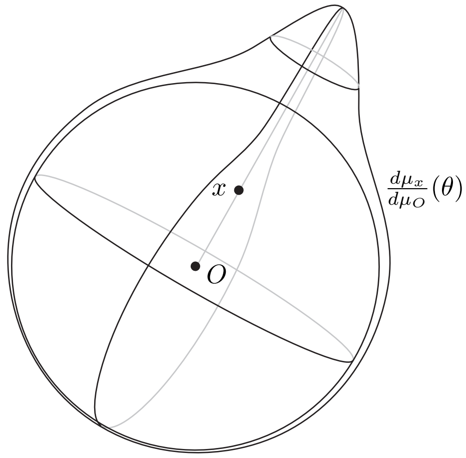


Figure 5.5: The Radon-Nikodym derivative of μ_x.

Proof.

1. Let A be a measurable subset of $\partial\mathbb{H}^n$. Let λ_x denote the canonical measure on UT_x. Since φ is an isometry, it follows from the definition of the canonical measure on the sphere that if B is a measurable subset of $UT_x\mathbb{H}^n$ then $\lambda_x(B) = \lambda_{\varphi(x)}(d_x\varphi(B))$.

   Then repeatedly applying definitions gives

   $$\mu_x(A) = \lambda_x(V_x(A)) = \lambda_{\varphi(x)}((d_x\varphi \circ V_x)(A)) = \lambda_{\varphi(x)}((V_{\varphi(x)} \circ \varphi)(A)) = \mu_{\varphi(x)}(\varphi(A))$$

   from which the result follows.

2. If $\varphi \in \mathrm{Isom}(\mathbb{H}^n)$ is such that $\varphi(O) = x$ then, by part 1, $\mu_x = \varphi_*[\mu_O]$. Since φ_* is a homeomorphism, μ_x is absolutely continuous with respect to μ_O.

3. We seek a function $p_x(\theta) : \partial\mathbb{H}^n \to \mathbb{R}$ such that for all measurable subsets $A \subseteq \partial\mathbb{H}^n$,

   $$\mu_x(A) = \int_{V_x(A)} d\lambda_x = \int_A p_x(\theta)\, d\mu_O(\theta).$$

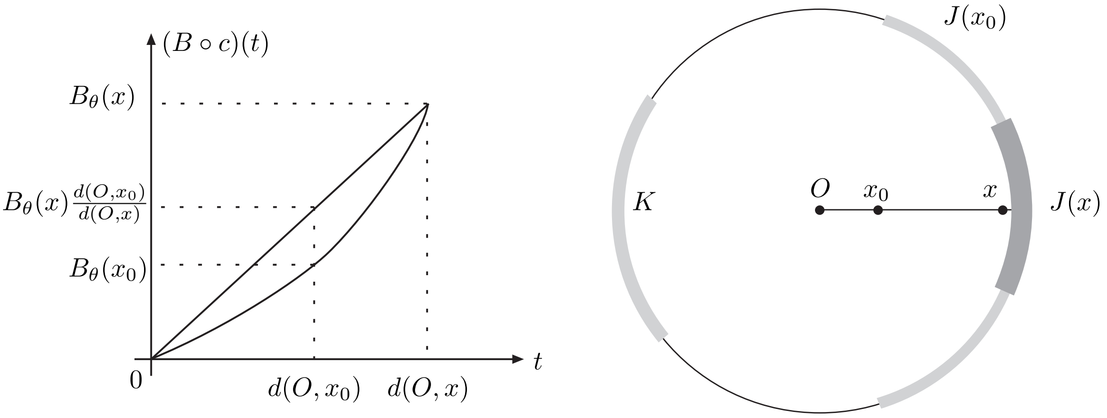
   Since λ_x and $d\mu_O$ are both the canonical measure on the sphere, it follows from the change of variables formula and Proposition 5.4 that

   $$p_x(\theta) = |\det(d_\theta V_x)| = e^{-(n-1)B(x,\theta)}. \qquad \square$$

5.3 The barycentre of a measure

In the previous section we showed how to assign to each $x \in \mathbb{H}^n$ a probability measure $\mu_x \in \mathcal{P}(\partial\mathbb{H}^n)$. In this section we describe a method of going back the other way — associating to (almost) any probability measure $\nu \in \mathcal{P}(\partial\mathbb{H}^n)$ a unique point $x \in \mathbb{H}^n$, known as the 'barycentre' of ν.

Given a probability measure $\nu \in \mathcal{P}(\partial\mathbb{H}^n)$ we want the barycentre of ν to be the point in $\mathbb{H}^n$ whose average 'distance' (with respect to ν) to the boundary $\partial\mathbb{H}^n$ is minimized. We make sense of this by minimizing an average (with respect to ν) of Busemann functions.



(a) Equation (5.1) illustrated.

(b)

Figure 5.6

As such let $\beta_\nu : \mathbb{H}^n \to \mathbb{R}$ be defined as

$$\beta_\nu(x) = \int_{\partial\mathbb{H}^n} B(x, \theta)\ d\nu(\theta).$$

Because Busemann functions are convex (see Section 5.1) it follows that β_ν is also convex. Indeed we can say more.

Proposition 5.7. *Given $\nu \in \mathcal{P}(\partial\mathbb{H}^n)$ the Hessian of β_ν is given by*

$$Hess_x(\beta_\nu)(u, u) = \langle u, u\rangle - \int_{\partial\mathbb{H}^n} \langle \nabla B_\theta(x), u\rangle^2\ d\nu(\theta).$$

As a consequence, if ν is atomless then β_ν is strictly convex.

Proof. Suppose $u \in T_x\mathbb{H}^n \setminus \{0\}$. Then since $\mathrm{Hess}_x(B_\theta)(u, u) \geq 0$,

$$\mathrm{Hess}_x(\beta_\nu)(u, u) = \int_{\partial\mathbb{H}^n} \mathrm{Hess}_x(B_\theta)(u, u)\ d\nu(\theta) \geq 0.$$

Equality occurs if and only if $\nabla B_\theta(x)$ is parallel to u for almost all θ in the support of ν, which is impossible if ν is atomless. So the inequality is strict and hence β_ν is strictly convex.

Using the fact that ν is a probability measure and the expression for $\mathrm{Hess}_x(B_\theta)(u, u)$ in Proposition 5.3, we obtain the required expression for $\mathrm{Hess}_x(\beta_\nu)(u, u)$. □

Definition 5.8. If ν is a probability measure on $\partial\mathbb{H}^n$ then a *barycentre* of ν is any $x \in \mathbb{H}^n$ that minimizes $\beta_\nu(x)$.

Proposition 5.9. *If $\nu \in \mathcal{P}(\partial\mathbb{H}^n)$ and ν is atomless then ν has a unique barycentre which we will denote $bar(\nu)$.*

Proof. (Following [BCG95, Appendix A].) Assuming that a barycentre exists, uniqueness follows from the strict convexity of β_ν when ν is atomless.

To show that the barycentre exists, we will show that $\beta_\nu(x) \to \infty$ as x moves along a radial path from O to $\partial\mathbb{H}^n$. It will then follow that β_ν achieves its minimum on $\mathbb{H}^n$.

For any $x \in \mathbb{H}^n$ let $J(x) = \{\theta \in \partial\mathbb{H}^n : B(x,\theta) \leq 0\}$. That is, $\theta \in J(x)$ if and only if x is inside the horoball passing through θ and O.

Suppose $c : [0,\infty) \to \mathbb{H}^n$ is a radial geodesic from O to $\theta_0 \in \partial\mathbb{H}^n$. (See Figure 5.6b). Choose $t_0 \in [0,\infty)$ large enough so that if $x_0 = c(t_0)$ then $\partial\mathbb{H}^n \setminus J(x_0)$ has strictly positive measure. Then let K be a compact subset of $\partial\mathbb{H}^n \setminus J(x_0)$ of strictly positive measure.

For any $t > t_0$ let $x = c(t)$. It follows from the convexity of B_θ and Figure 5.6a that

$$B_\theta(x_0) \leq \frac{B_\theta(x)}{d(x,O)} \cdot d(x_0,O). \tag{5.1}$$

Furthermore, since $t > t_0$, $J(x) \subset J(x_0)$ so

$$\begin{aligned}
\beta_\nu(x) &= \int_{J(x)} B(x,\theta)\, d\nu(\theta) + \int_{\partial\mathbb{H}^n \setminus J(x)} B(x,\theta)\, d\nu(\theta) \\
&\geq \int_{J(x)} B(x,\theta)\, d\nu(\theta) + \int_K B(x,\theta)\, d\nu(\theta) \\
&\geq \frac{d(O,x)}{d(O,x_0)} \int_{J(x)} B(x_0,\theta)\, d\nu(\theta) + \frac{d(O,x)}{d(O,x_0)} \int_K B(x_0,\theta)\, d\nu(\theta) \\
&\geq \frac{d(O,x)}{d(O,x_0)} \left[\min_{\theta\in\partial\mathbb{H}^n} \{B(x_0,\theta)\}\nu(J(x)) + \max_{\theta\in K}\{B(x_0,\theta)\} \right]. \qquad (5.2)
\end{aligned}$$

Notice that the expression in the brackets in (5.2) is the sum of a negative term (depending on x) and a constant positive term. As $t \to \infty$ we see that $\bigcap_{t\geq t_0} J(x(t)) = \{\theta_0\}$. Since ν is atomless, $\nu(J(c(t))) \to \nu(\{\theta_0\}) = 0$. Hence the right hand side of (5.2) goes to infinity as $t \to \infty$. ☐

Since the barycentre is defined as the minimum of the differentiable function β_ν, $\mathrm{bar}\,(\nu)$ must be a critical point of β_ν. Hence if $x = \mathrm{bar}\,(\nu)$ then

$$\nabla\beta_\nu(x) = \int_{\partial\mathbb{H}^n} \nabla B_\theta(x)\, d\nu(\theta) = 0 \in T_x\mathbb{H}^n. \tag{5.3}$$

This implicit equation will play a crucial role later on. Equivalently if $x = \mathrm{bar}\,(\nu)$ then the implicit equation can be written

$$\int_{\partial\mathbb{H}^n} d_x B_\theta(\cdot)\, d\nu(\theta) = 0(\cdot)$$

which is to be understood as an equality of linear functionals.

The final property of the barycentre that we will need is that it is equivariant with respect to isometries.

Proposition 5.10. *If ν is an atomless probability measure on $\partial\mathbb{H}^n$ and $\varphi \in Isom(\mathbb{H}^n)$ then*

$$\mathit{bar}\,(\varphi_*[\nu]) = \varphi(\mathit{bar}\,(\nu)).$$

Proof. Let $x = \varphi(\mathrm{bar}\,(\nu))$ and let $y = \varphi^{-1}(x) = \mathrm{bar}\,(\nu)$. Then the implicit equation for the barycentre tells us that

$$\begin{aligned}
0 = \int_{\partial\mathbb{H}^n} d_y B_\theta(\cdot)\, d\nu(\theta) &= \int_{\partial\mathbb{H}^n} d_x B_{\varphi(\theta)}(d_y\varphi(\cdot))\, d\nu(\theta) \\
&= \left(\int_{\partial\mathbb{H}^n} d_x B_\theta(\cdot)\, d\varphi_*[\nu](\theta) \right) \circ d_y\varphi.
\end{aligned}$$

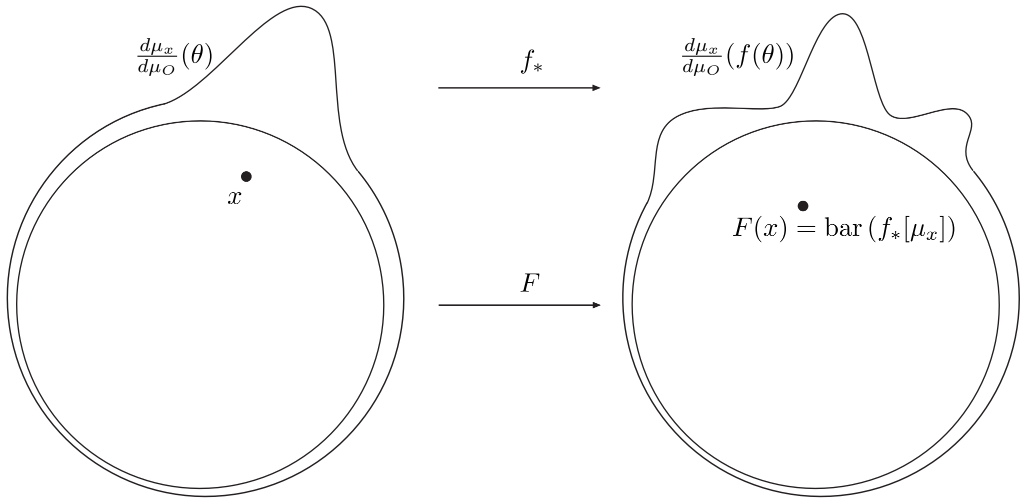


Figure 5.7: The barycentric extension of f.

Since $d_y\varphi$ is invertible,

$$\int_{\partial\mathbb{H}^n} d_x B_\theta(\cdot)\; d\varphi_*[\nu](\theta) = 0$$

so by uniqueness of solutions to the implicit equation, $x = \mathrm{bar}\,(\varphi_*[\nu])$. $\square$

5.4 The barycentric extension

Suppose $f : \partial\mathbb{H}^n \to \partial\mathbb{H}^n$ is a homeomorphism. In this section we explain one method to extend f to a particularly nice map $F : \mathbb{H}^n \to \mathbb{H}^n$.

Definition 5.11. If $f : \partial\mathbb{H}^n \to \partial\mathbb{H}^n$ is a homeomorphism then the *barycentric extension of* f is a map $F : \mathbb{H}^n \to \mathbb{H}^n$ defined by

$$F(x) = \mathrm{bar}\,(f_*[\mu_x])$$

where μ_x is the visual measure associated with x.

We have seen in Proposition 5.6 that μ_x is atomless for any $x \in \mathbb{H}^n$. Since f is a homeomorphism it follows that $f_*[\mu_x]$ is atomless for all x. Thus $\mathrm{bar}\,(f_*[\mu_x])$ exists and is unique for all $x \in \mathbb{H}^n$ and so F is well defined.

Since the barycentre of an atomless measure satisfies an implicit equation, F also satisfies an implicit equation. To simplify notation a little, we will write β_x instead of $\beta_{f_*[\mu_x]}$ to denote the function

$$\beta_x(y) = \int_{\partial\mathbb{H}^n} B_\theta(y)\; df_*[\mu_x](\theta).$$

Then applying the implicit equation for the barycentre of $f_*[\mu_x]$ we see that the barycentric extension satisfies

$$\nabla\beta_x(F(x)) = \int_{\partial\mathbb{H}^n} \nabla B_\theta(F(x))\; df_*[\mu_x](\theta) = 0 \in T_{F(x)}\mathbb{H}^n \tag{5.4}$$

for all $x \in \mathbb{H}^n$.

The remainder of this section is devoted to establishing various properties of F.

Lemma 5.12. *If $f : \partial\mathbb{H}^n \to \partial\mathbb{H}^n$ is a homeomorphism with barycentric extension F then F is smooth.*

Proof. Since F satisfies an implicit equation, we will use the implicit function theorem to show that F is smooth.

Fix a point $x_0 \in \mathbb{H}^n$ and let $y_0 = F(x_0)$. Choose an orthonormal basis $\{E_1(y_0), E_2(y_0), \ldots, E_n(y_0)\}$ for $T_{y_0}\mathbb{H}^n$. Define an orthonormal basis $\{E_i(y)\}$ at any $y \in \mathbb{H}^n$ by parallel transport along the unique geodesic segment joining y_0 and y.

Then define a function $G : \mathbb{H}^n \times \mathbb{H}^n \to \mathbb{R}^n$ whose component functions are

$$G_i(x, y) = \langle \nabla\beta_x(y), E_i(y)\rangle = \int_{\partial\mathbb{H}^n} \langle \nabla B_\theta(y), E_i(y)\rangle \; df_*[\mu_x](\theta).$$

Since $\nabla B_\theta(y)$ is a smooth function, it follows that G is also smooth.

For some $\epsilon > 0$ let $\gamma : (-\epsilon, \epsilon) \to \mathbb{H}^n$ be a geodesic segment with $\gamma(0) = y_0$ and $\gamma'(0) = u$. Then

$$\begin{aligned} \frac{d}{dt} G_i(x, \gamma(t))\bigg|_{t=0} &= \frac{d}{dt} \langle \nabla\beta_x(\gamma(t)), E_i(\gamma(t))\rangle\bigg|_{t=0} \\ &= \langle D_u\nabla\beta_x(y_0), E_i(y_0)\rangle + \langle \nabla\beta_x(y_0), D_u E_i(y_0)\rangle \\ &= \mathrm{Hess}_{y_0}(\beta_x)(u, E_i(y_0)) \end{aligned}$$

(where the last equality holds because $E_i(\gamma(t))$ is parallel along γ). It follows from Proposition 5.7 that $\mathrm{Hess}_{y_0}(\beta_x)$ is positive definite. So for all non-zero $u \in T_{y_0}\mathbb{H}^n$,

$$\frac{d}{dt} G_i(x, \gamma(t))\bigg|_{t=0} > 0.$$

Hence the derivative of each $G_i(x, y)$ with respect to the second variable has trivial kernel, so the same holds for the derivative of G with respect to the second variable.

Since G is smooth and the derivative of G with respect to the second variable is invertible, all the conditions of the implicit function theorem are satisfied. So it follows that F agrees with a smooth function on a neighbourhood of x_0. Since x_0 was arbitrary, the proof is complete. $\square$

The next result establishes the volume decreasing properties that make the barycentric extension such a remarkable construction. The proof is quite involved, but it will be worth the effort. A proof of Mostow's rigidity theorem comes as a fairly straightforward corollary of this result.

The main object of interest here is the Jacobian of the barycentric extension F. This is, roughly, the volume scaling factor of F at each point. One way to compute this is to choose orthonormal bases for $T_x\mathbb{H}^n$ and $T_{F(x)}\mathbb{H}^n$ and find the determinant of the matrix of the derivative of F written with respect to these bases. Another way to compute the Jacobian, the way we adopt here, is to compute $\sqrt{|det((d_xF)^* d_xF)|}$ where $(d_xF)^*$ is the adjoint of d_xF. Note that since $T_x\mathbb{H}^n$ and $T_{F(x)}\mathbb{H}^n$ have the same dimension, $\sqrt{|det((d_xF)^* d_xF)|} = \sqrt{|det(d_xF(d_xF)^*)|}$, a fact we will exploit later on.

Proposition 5.13. *Suppose $n \geq 3$. If $f : \partial\mathbb{H}^n \to \partial\mathbb{H}^n$ is a homeomorphism and F is its barycentric extension then*

$$|Jac_x(F)| := \sqrt{|\det((d_xF)^* d_xF)|} \leq 1$$

for all $x \in \mathbb{H}^n$. Furthermore, if $|Jac_x(F)| = 1$ then $d_xF : T_x\mathbb{H}^n \to T_{F(x)}\mathbb{H}^n$ is an isometry.

Proof. The proof we give largely follows the argument of Besson et al. [BCG96] as presented by Pansu [Pan97]. Before delving into the details, we give a brief outline of the proof.

Differentiating the implicit equation gives rise to an expression for the derivative of F. After some simplification (which will take considerable effort), we can bound $dF(dF)^*$ (in terms of the usual partial order on positive operators) by a fairly simple combination of positive definite linear endomorphisms.

Applying an elementary (yet non-trivial) inequality related to the determinants of positive definite matrices with trace one (Lemma 5.17), we can conclude that F is volume non-increasing, and preserves volume at x if and only if d_xF is an isometry.

Throughout, fix $x \in \mathbb{H}^n$. Let $u \in T_x\mathbb{H}^n$ and, for some small $\epsilon > 0$ let $\gamma : (-\epsilon, \epsilon)$ be a geodesic segment with $\gamma(0) = x$ and $\gamma'(0) = u$. Let $p(x) = e^{-(n-1)B_\theta(x)}$ be the Radon-Nikodym derivative of μ_x with respect to μ_O. Then it follows from the implicit equation for F (5.4) that

$$\int_{\partial\mathbb{H}^n} p(\gamma(t))\nabla B_{f(\theta)}((F \circ \gamma)(t))\; d\mu_O(\theta) = 0 \tag{5.5}$$

for all $t \in (-\epsilon, \epsilon)$. Observe that $V(t) := p(\gamma(t))\nabla B_{f(\theta)}((F \circ \gamma)(t))$ is a vector field along the curve $F \circ \gamma$. Since the velocity vector of the curve is $d_xF(u)$, taking the covariant derivative of $V(t)$ and applying the Leibniz rule gives

$$\frac{DV}{dt} = d_xp(u)\nabla B_{f(\theta)}(F(x)) + p(x)D_{d_xF(u)}\nabla B_{f(\theta)}(F(x)).$$

Taking the inner product of both sides with some $v \in T_{F(x)}\mathbb{H}^n$ yields

$$\left\langle \frac{DV}{dt}, v \right\rangle = d_xp(u)d_{F(x)}B_{f(\theta)}(v) + p(x)\mathrm{Hess}_{F(x)}(B_{f(\theta)})(d_xF(u), v). \tag{5.6}$$

Differentiating under the integral sign in (5.5), substituting in the expression in (5.6), and noting that $d_xp(\cdot) = -(n-1)d_xB_\theta(\cdot)p(x)$, we obtain

$$(n-1)\int_{\partial\mathbb{H}^n} d_xB_\theta(u)d_{F(x)}B_{f(\theta)}(v)\; d\mu_x(\theta) = \int_{\partial\mathbb{H}^n} \mathrm{Hess}_{F(x)}(B_{f(\theta)})(d_xF(u), v)\; d\mu_x(\theta).$$

Replacing $\mathrm{Hess}_{F(x)}(B_{f(\theta)})(d_xF(u), v)$ with the expression we derived in Section 5.1 we obtain

$$\begin{aligned}(n-1)\int_{\partial\mathbb{H}^n} d_xB_\theta(u)d_{F(x)}B_{f(\theta)}(v)\; d\mu_x(\theta) &= \\ \langle d_xF(u), v\rangle - \int_{\partial\mathbb{H}^n} d_{F(x)}B_{f(\theta)}(d_xF(u))d_{F(x)}B_{f(\theta)}(v)\; d\mu_x&.\end{aligned} \tag{5.7}$$

Now define $\phi : T_x\mathbb{H}^n \to L^2(\partial\mathbb{H}^n, \mu_x)$ and $\psi : T_{F(x)}\mathbb{H}^n \to L^2(\partial\mathbb{H}^n, \mu_x)$ by

$$\phi(u) = (\theta \mapsto \sqrt{n}d_xB_\theta(u)) \quad \text{and} \quad \psi(v) = (\theta \mapsto \sqrt{n}d_{F(x)}B_{f(\theta)}(v)).$$

These are linear maps between inner product spaces. In what follows we frequently make use of the adjoints of these maps, which we denote $\phi^* : L^2(\partial\mathbb{H}^n, \mu_x) \to T_x\mathbb{H}^n$ and $\psi^* : L^2(\partial\mathbb{H}^n, \mu_x) \to T_{F(x)}\mathbb{H}^n$.

We can now rewrite (5.7) much more cleanly as

$$\langle d_xF(u), v\rangle - \frac{1}{n}\langle \psi(d_xF(u)), \psi(v)\rangle_{L^2} = \frac{n-1}{n}\langle \phi(u), \psi(v)\rangle_{L^2}.$$

By taking adjoints and noting that $u \in T_x\mathbb{H}^n$ and $v \in T_{F(x)}\mathbb{H}^n$ were arbitrary, we finally get an expression for the derivative of the barycentric extension:

$$(I - \frac{1}{n}\psi^*\psi)d_xF = \frac{n-1}{n}\psi^*\phi. \tag{5.8}$$

From here on, if we have two self-adjoint operators A and B we will write $A \leq B$ to mean that $B - A$ is positive semi-definite.

We will defer the proofs of the following properties of ϕ and ψ until Section 5.7.

Lemma 5.14. $tr\left(\frac{1}{n}\psi^*\psi\right) = 1$

Lemma 5.15. *φ is an isometric embedding of $T_x\mathbb{H}^n$ into $L^2(\partial\mathbb{H}^n, \mu_x)$. That is $\phi^*\phi = I$. Furthermore, $\phi\phi^* \leq I$.*

The next result will allow us to translate from the ordering on self-adjoint operators to ordering the traces and determinants of such operators. Again we defer the proof until Section 5.7.

Lemma 5.16. *If A and B are self-adjoint matrices and $A \leq B$ then $tr(A) \leq tr(B)$. If, in addition, A is positive definite then* $\det(A) \leq \det(B)$.

Now let $H = \frac{1}{n}\psi^*\psi$. It is clear that H is positive definite and we have shown in Lemma 5.14 that H has trace one. Hence all the eigenvalues of H lie in $(0, 1)$ so $I - H$ is positive definite. Applying Lemma 5.15 gives

$$\frac{n}{(n-1)^2} d_xF(d_xF)^* = \frac{1}{n}(I-H)^{-1}\psi^*\phi\phi^*\psi(I-H)^{-1} \leq (I-H)^{-1}H(I-H)^{-1}. \tag{5.9}$$

The proof of the entire proposition now hinges on the following rather mysterious inequality. Note that the inequality fails spectacularly for $n = 2$. This is the only place where we require $n \geq 3$ in this argument.

Lemma 5.17. *Suppose $n \geq 3$. If H is an $n \times n$ symmetric positive definite matrix of trace one then*

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

with equality if and only if $H = \frac{1}{n}I$.

The proof is elementary and yet rather involved so we defer it until Section 5.7.

By taking the determinant of both sides of (5.9), and applying Lemma 5.17 to the right hand side we see that

$$|\mathrm{Jac}_x(F)| = \sqrt{\det(d_xF(d_xF)^*)} \leq \sqrt{\left(\frac{(n-1)^2}{n}\right)^n \left(\frac{n}{(n-1)^2}\right)^n} = 1 \tag{5.10}$$

with equality if and only if $H = \frac{1}{n}I$.

Consider the equality case. To proceed we need the following standard inequality.

Lemma 5.18. *If H is an $n \times n$ symmetric positive definite matrix then*

$$\det(H) \leq \left(\frac{tr(H)}{n}\right)^n$$

with equality if and only if $H = (\det(H))^{1/n}I$.

Proof. Apply the AM-GM inequality to the eigenvalues of H, all of which are positive real numbers. $\square$

In the equality case of (5.10), $\det(d_xF(d_xF)^*) = 1$, $H = \frac{1}{n}I$, and $(I-H)^{-1} = \frac{n-1}{n}I$. So

$$\begin{aligned} 1 &= \det(d_xF(d_xF)^*) \leq \left(\frac{\operatorname{tr}(d_xF(d_xF)^*)}{n}\right)^n \\ &\leq \left[\operatorname{tr}\left(\left(\frac{n-1}{n}\right)^2 (I-H)^{-1}H(I-H)^{-1}\right)\right]^n \\ &= \operatorname{tr}(H)^n = 1. \end{aligned}$$

Hence we must be in the equality case of Lemma 5.18. Consequently

$$d_xF(d_xF)^* = (\det(d_xF(d_xF)^*))^{1/n}I = I$$

so d_xF is an isometry. $\square$

5.5 A proof of Mostow's theorem

In this section we finally prove the following.

Theorem 5.19. *Suppose $n \geq 3$ and $M_1 = \mathbb{H}^n/\Gamma_1$ and $M_2 = \mathbb{H}^n/\Gamma_2$ are closed orientable hyperbolic n-manifolds. If $\psi : \Gamma_1 \to \Gamma_2$ is an isomorphism then φ is induced by an isometry $F : M_1 \to M_2$.*

Proof. In Chapter 4 we showed that given an isomorphism $\varphi : M_1 \to M_2$ there is a homeomorphism

$$\partial f : \partial\mathbb{H}^n \to \partial\mathbb{H}^n$$

such that for any $\gamma \in \Gamma_1$

$$\partial f \circ \gamma = \psi(\gamma) \circ \partial f. \tag{5.11}$$

Let $\tilde{F}$ be the barycentric extension of ∂f. Then if $\gamma \in \Gamma_1$ it follows from (5.11) and the equivariance properties of the visual measure and the barycentre (Propositions 5.6 and 5.10) that for all $x \in \mathbb{H}^n$,

$$\begin{aligned} (\tilde{F} \circ \gamma)(x) &= \operatorname{bar}\left(\partial f_*[\mu_{\gamma(x)}]\right) \\ &= \operatorname{bar}\left((\partial f \circ \gamma)_*[\mu_x]\right) \\ &= \operatorname{bar}\left((\psi(\gamma))_*[\partial f_*[\mu_x]]\right) \\ &= \psi(\gamma)(\operatorname{bar}(\partial f_*[\mu_x])) \\ &= (\psi(\gamma) \circ \tilde{F})(x). \end{aligned}$$

So $\tilde{F}$ descends to a map $F : M_1 \to M_2$ that induces an isomorphism on the fundamental groups of M_1 and M_2. Then by the corollary of Whitehead's theorem for hyperbolic manifolds (Corollary 2.14), F is actually a homotopy equivalence and so is a map of degree ± 1 depending on whether F preserves or reverses orientation.

Let Ω_1 and Ω_2 denote the volume forms on M_1 and M_2 respectively. Let $[M_1]$ and $[M_2]$ denote fundamental classes of M_1 and M_2 respectively and let F_* denote the map on homology incuded by F. Then, by the definition of the degree of a map,

$$F_*([M_1]) = \deg(F)[M_2] = \pm[M_2].$$

So if $\langle \cdot, \cdot \rangle$ denotes the pairing between de Rham cohomology classes and singular homology classes induced by integration, and $F^* : TM_2^* \to TM_1^*$ is the pullback of F, then

$$\mathrm{Vol}(M_2) = \langle [\Omega_2], [M_2] \rangle = \langle [\Omega_2], \pm F_*([M_1]) \rangle = \pm \langle [F^*(\Omega_2)], [M_1] \rangle = \pm \int_{M_1} F^*(\Omega_2).$$

If F is orientation preserving then, by Proposition 5.13,

$$\mathrm{Vol}(M_2) = \int_{M_1} F^*(\Omega_2) = \int_{M_1} |\mathrm{Jac}(F)| \Omega_1 \le \mathrm{Vol}(M_1).$$

If F is orientation reversing then, if $\overline{M_1}$ denotes M_1 with the opposite orientation then

$$\mathrm{Vol}(M_2) = -\int_{M_1} F^*(\Omega_2) = -\int_{\overline{M_1}} |\mathrm{Jac}(F)| \Omega_1 = \int_{M_1} |\mathrm{Jac}(F)| \Omega_1 \le \mathrm{Vol}(M_1).$$

By reversing the roles of M_1 and M_2 and performing the same construction we find that

$$\mathrm{Vol}(M_1) \le \mathrm{Vol}(M_2).$$

Hence $\mathrm{Vol}(M_1) = \mathrm{Vol}(M_2)$. In this case we must have $|\mathrm{Jac}_x(F)| = 1$ for almost all $x \in M_1$. But then we are in the equality case of Proposition 5.13 so $d_x F$ is an isometry for almost all x. By continuity this holds for all $x \in \mathbb{H}^n$ so F is an isometry. □

5.6 The bigger picture: volume, entropy, and rigidity

Besson, Courtois, and Gallot's proof of Mostow Rigidity arises as a corollary of a much more general result. In this section we briefly outline this more general setting and discuss the extent to which the arguments of the first part of this chapter carry through to this setting.

5.6.1 Volume entropy

If (Y, g) is a compact connected Riemannian n-manifold, let $(\tilde{Y}, \tilde{g})$ be the universal cover of Y with the pulled-back metric. Then for any $y \in \tilde{Y}$ define the *volume entropy* to be

$$h(g) = \lim_{R \to \infty} \frac{1}{R} \log \left(\mathrm{vol}(B(y, R)) \right)$$

where $B(y, R)$ is the ball of radius R centred at $y \in \tilde{Y}$. It turns out that this limit exists and is independent of the choice of y. Essentially it measures the growth rate of balls in the universal cover of Y. For example, if (Y, g) is a hyperbolic n-manifold, $(\tilde{Y}, \tilde{g})$ is isometric to $\mathbb{H}^n$. In this case we know that $\mathrm{vol}(B(y, R)) \sim e^{(n-1)R}$ from our discussion of the volume of balls in $\mathbb{H}^n$ in Section 2.1.3. Hence

$$h(g) = \lim_{R \to \infty} \frac{1}{R} \log(e^{(n-1)R}) = n - 1.$$

5.6.2 Entropy and Volume characterize 'nice' metrics

Imagine we are given a hyperbolic manifold (X, g_0) and some other Riemannian manifold (Y, g). Suppose, further, that there is a continuous map $f : Y \to X$ of non-zero degree. Amazingly, it turns out that just by calculating the volume and volume entropy of X and Y and comparing these numbers in the right way, we can tell whether (Y, g) is also hyperbolic.

More precisely, Besson et al. prove the following in [BCG95].

Proposition 5.20. *For $n \geq 3$, let (Y, g) be a closed orientable Riemannian n-manifold. Let (X, g_0) be an n-dimensional closed hyperbolic manifold. If $f : Y \to X$ is a continuous map of non-zero degree then*

$$h(g)^n \, Vol(Y, g) \geq |deg(f)| h(g_0)^n \, Vol(X, g_0).$$

Equality holds if and only if f is homotopic to a covering map that is a local isometry (after rescaling g by an appropriate constant).

If we set $\deg(f) = 1$ and $h(g) = n - 1$ then this statement looks rather similar to the expressions we obtained in section 5.5 on our way to proving Mostow's Rigidity Theorem.

Before, we assumed that both spaces were hyperbolic. Now, no such assumption is made on (Y, g). Yet if (Y, g) has a metric of strictly negative curvature and f is a homotopy equivalence then the general outline of the proof we gave still works, as long as a few modifications are made.

First, while the definition of the Busemann function on Y still makes sense, we can no longer be sure that it is smooth. It does turn out to be C^2, though [BCG95]. Hence the implicit function theorem will still tell us that the barycentric extension is C^1. Second, we can no longer use the 'visual' definition of the measures μ_x that we use to embed $\tilde{Y}$ into the space of probability measures on $\partial\tilde{Y}$. It turns out that in this more general context, the so-called Patterson-Sullivan measures appropriately generalize the role of the visual measures. (See [Nic91] for an introduction to this remarkable family of measures.) Finally, the map $\phi : T_x\tilde{Y} \to L^2(\partial\tilde{Y}, \mu_x)$ which played a role in our proof that the barycentric extension map is volume non-increasing, no longer satisfies $\phi^*\phi = I$. It turns out that this is not such a problem — the same bounds on the Jacobian can be obtained by other, slightly less elegant, means.

In the more general case, that is when we have *any* Riemannian metric on Y, the technique of pushing everything out the the boundary 'at infinity' no longer works, as such a boundary does not enjoy the same nice properties as in the negatively-curved case. Nevertheless in [BCG95] the authors still manage to deal with this case, albeit with a less directly constructive argument.

Besson, Courtois, and Gallot's results are actually more general than the result we have introduced in this section. Instead of just being able to characterize hyperbolic metrics, it turns out that the volume and volume entropy together conspire to characterize the negatively curved locally symmetric metrics on Riemannian manifolds. The modifications required to make this generalization are surprisingly minor. The main thing that changes is that the expressions for the Hessian of the Busemann function becomes more complicated, and a stronger version of the 'mysterious' inequality (Lemma 5.17) is required to deal with this situation.

5.6.3 Extension to the finite volume case

Storm [Sto06] has recently proved a result analogous to Proposition 5.20 in the case where the manifolds have finite volume.[1]

Proposition 5.21. *For $n \geq 3$, let (Y, g) is a finite volume orientable Riemannian n-manifold. Let (X, g_0) be a finite volume orientable hyperbolic n-manifold. If $f : Y \to X$ is a continuous map of non-zero degree then*

$$h(g)^n \, Vol(Y, g) \geq |deg(f)| h(g_0)^n \, Vol(X, g_0).$$

Equality holds if and only if f is homotopic to a covering map that is a local isometry (after rescaling g by an appropriate constant).

[1] Again this result, like the results of Besson et al. also hold in the locally symmetric case.

Note that given this proposition we can easily give a proof of Mostow rigidity in the finite volume case — the case first proved by Prasad in [Pra73] and sometimes known as Mostow-Prasad rigidity.

Theorem 5.22. *For $n \geq 3$, let (Y, g) and (X, g_0) be finite volume orientable hyperbolic n-manifolds. If $f : Y \to X$ is a homotopy equivalence then f is an isometry.*

Proof. Since f is a homotopy equivalence $\deg(f) = 1$. Since (Y, g) is hyperbolic, $h(g) = h(g_0) = n - 1$. Applying Proposition 5.21 we have $\mathrm{Vol}(Y, g) \geq \mathrm{Vol}(X, g_0)$. Reversing the roles of X and Y gives the reverse inequality. So we must be in the equality case of Proposition 5.21. But then f is homotopic to a local isometry of degree one, i.e. an isometry. $\square$

5.7 Details

In this section we give proofs of a number of lemmas from Section 5.4.

Proof of lemma 5.14. Let $e_1, \ldots, e_n$ be an orthonormal basis for $T_{F(x)}\mathbb{H}^n$. Then for any $\theta \in \partial\mathbb{H}^n$,

$$\nabla B_{f(\theta)}(F(x)) = \sum_{i=1}^{n} \left\langle \nabla B_{f(\theta)}(F(x)), e_i \right\rangle e_i.$$

Since the gradient of the Busemann function is a unit vector,

$$1 = \|\nabla B_{f(\theta)}(F(x))\|^2 = \sum_{i=1}^{n} \left\langle \nabla B_{f(\theta)}(F(x)), e_i \right\rangle^2.$$

Hence

$$\begin{aligned}
\mathrm{tr}\left(\frac{1}{n}\psi^*\psi\right) = \frac{1}{n}\sum_{i=1}^{n} \langle \psi^*\psi(e_i), e_i \rangle &= \frac{1}{n}\sum_{i=1}^{n} \langle \psi(e_i), \psi(e_i) \rangle_{L^2} \\
&= \frac{1}{n}\sum_{i=1}^{n} \int_{\partial\mathbb{H}^n} [\sqrt{n} d_{F(x)} B_{f(\theta)}(e_i)]^2 \, d\mu_x \\
&= \int_{\partial\mathbb{H}^n} \sum_{i=1}^{n} \left\langle \nabla B_{f(\theta)}(F(x)), e_i \right\rangle^2 \, d\mu_x \\
&= \int_{\partial\mathbb{H}^n} d\mu_x = 1
\end{aligned}$$

since μ_x is a probability measure. $\square$

Proof of lemma 5.15. To show that $\phi^*\phi = I$ we can, equivalently, show that for all unit vectors $u \in T_x\mathbb{H}^n$, $\langle \phi(u), \phi(u) \rangle_{L^2} = \langle u, u \rangle = 1$.

Let $u \in T_x\mathbb{H}^n$ and $v \in T_O\mathbb{H}^n$ be unit vectors. Let φ be an isometry of $\mathbb{H}^n$ such that $\varphi(x) = O$ and $d_x\varphi(u) = v$. Then using properties of the Busemann function from Proposition 5.3 and properties of the visual measure from Proposition 5.6 we obtain

$$\begin{aligned}
\langle \phi(u), \phi(u) \rangle_{L^2} &= n \int_{\partial \mathbb{H}^n} [d_x B_\theta(u)]^2 \, d\mu_x(\theta) \\
&= n \int_{\partial \mathbb{H}^n} [d_{\varphi(x)} B_{\varphi(\theta)}(d_x \varphi(u))]^2 \, d\mu_x(\theta) \\
&= n \int_{\partial \mathbb{H}^n} [d_O B_\theta(v)]^2 \, d\varphi_*[\mu_x](\theta) \\
&= n \int_{\partial \mathbb{H}^n} [d_O B_\theta(v)]^2 \, d\mu_O(\theta) = \langle \phi(v), \phi(v) \rangle_{L^2}.
\end{aligned}$$

So it suffices to show that $\langle \phi(v), \phi(v) \rangle_{L^2} = 1$ for one unit vector $v \in T_O \mathbb{H}^n$. This is particularly easy because in this case μ_O is the canonical measure on the sphere.

Let us move to spherical coordinates $(\theta_1, \dots, \theta_{n-2}, \theta_{n-1})$ (see Section 2.4 of Ratcliffe's book [Rat94], for example, for a precise description of the notation and derivation of the associated volume form). Take $v = e_1$ to be the unit vector in the first coordinate direction. Then the integrand is exactly $\cos^2(\theta_1)$. So we have

$$\frac{1}{n} \langle \phi(v), \phi(v) \rangle_{L^2} = \frac{\int_0^\pi \cdots \int_0^\pi \int_0^{2\pi} \cos^2(\theta_1) \sin^{n-2}(\theta_1) \cdots \sin(\theta_{n-2}) \, d\theta_1 \cdots d\theta_{n-2} d\theta_{n-1}}{\int_0^\pi \cdots \int_0^\pi \int_0^{2\pi} \sin^{n-2}(\theta_1) \cdots \sin(\theta_{n-2}) \, d\theta_1 \cdots d\theta_{n-2} d\theta_{n-1}}$$

where we have normalized the volume form to have total volume one. This can be simplified to

$$\frac{1}{n} \langle \phi(v), \phi(v) \rangle_{L^2} = \frac{\int_0^\pi (1 - \sin^2(\theta_1)) \sin^{n-2}(\theta_1) \, d\theta_1}{\int_0^\pi \sin^{n-2}(\theta_1) \, d\theta_1} = 1 - \frac{\int_0^\pi \sin^n(\theta) \, d\theta}{\int_0^\pi \sin^{n-2}(\theta) \, d\theta}.$$

Integrating $\sin^n(\theta) = \frac{1}{n} n \sin^{n-1}(\theta) \sin(\theta)$ by parts and rearranging things gives

$$\int_0^\pi \sin^n \theta \, d\theta = \frac{n-1}{n} \int_0^\pi \sin^{n-2} \theta \, d\theta.$$

From this it follows that

$$\langle \phi(v), \phi(v) \rangle_{L^2} = n \left(1 - \frac{n-1}{n} \right) = 1.$$

So $\phi^* \phi = I$.

To address the second statement in the lemma, it is enough to observe that since $\phi^* \phi = I$ it follows that $(\phi \phi^*)^2 = \phi \phi^*$. Because, in addition, $\phi \phi^*$ is self-adjoint, $\phi \phi^*$ is the orthogonal projection onto its image. It follows immediately that $\phi \phi^* \leq I$. □

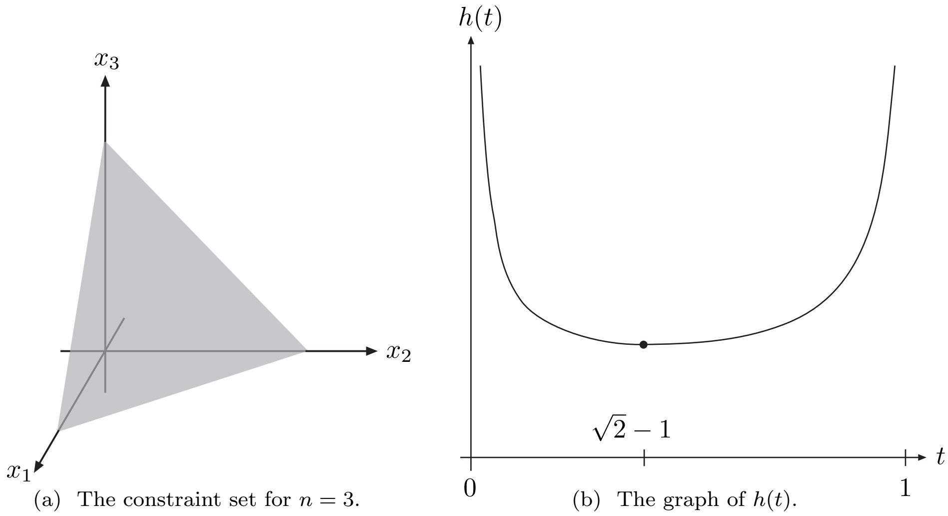
Proof of lemma 5.16. If $A \leq B$ then $B - A$ is positive semi-definite so has all non-negative eigenvalues. Hence $\mathrm{tr}(B - A) \geq 0$. Since the trace is a linear operation it follows that $\mathrm{tr}(A) \leq \mathrm{tr}(B)$.

If A is also positive definite then we can write $A = PP^*$ for some positive definite matrix P. Let $Q = (P^*)^{-1}$ so that $Q^* A Q = I$. Observe that because B is self-adjoint, so is $Q^* B Q$. Hence there is some unitary U and some diagonal D such that $(QU)^* B (QU) = D$. Furthermore, $(QU)^* A (QU) = U^* Q^* A Q U = I$.

Now if $A \leq B$ then $(QU)^* A (QU) \leq (QU)^* B (QU)$. Equivalently $I \leq D$. Then it follows that $1 = \det(I) \leq \det(D)$. Then

$$\det((QU)^* QU) \det(A) \leq \det((QU)^* QU) \det(B).$$

Since $\det((QU)^* QU) = \det(Q^* Q) > 0$ the result follows. □



(a) The constraint set for $n = 3$. (b) The graph of $h(t)$.

Figure 5.8

Finally we give a proof of the inequality that was at the heart of the proof that the barycentric extension is volume non-increasing. The proof we give, which follows the proof in Appendix B of [BCG95], is rather computational and unenlightening. To our knowledge, no conceptual proof of this result is known. Such a proof may shed more light on why the techniques of Besson, Courtois, and Gallot are as successful as they are.

Proof of lemma 5.17. Since H is symmetric and positive definite with trace one its eigenvalues $x_1, \ldots, x_n$ are positive and sum to one. Then the statement of the lemma is equivalent to showing that if $x \in \operatorname{int}(\Delta^{n-1})$, the interior of the standard $(n-1)$-simplex, then

$$f(x) = \frac{\prod_{i=1}^n x_i}{\prod_{i=1}^n (1-x_i)^2} \leq \left(\frac{n}{(n-1)^2}\right)^n$$

with equality if and only if each $x_i = 1/n$.

Taking logarithms, we will show that $g(x) = \log(f(x)) = \sum_i \ln(x_i) - 2\sum_i \ln(1-x_i)$ has exactly one critical point on the interior of the standard $(n-1)$-simplex, at $x = (1/n, 1/n, \ldots, 1/n)$. Then, checking that g does not achieve a maximum on the boundary of the simplex will prove the result.

Using Lagrange mutipliers, any critical point $(x_1, \ldots, x_n)$ in the interior of the standard $(n-1)$-simplex must satisfy $\nabla g(x) = \lambda(1, 1, \ldots, 1)$ for some $\lambda \in \mathbb{R}$. Equivalently,

$$\frac{1}{x_i} + \frac{2}{1-x_i} = \frac{1}{x_j} + \frac{2}{1-x_j} \tag{5.12}$$

for all $1 \leq i, j \leq n$. Note that if $x_1 = x_2 = \cdots = x_n = 1/n$ then this is certainly satisfied.

Let $h(t) = 1/t + 2/(1-t)$ and observe that $h(t)$ has a unique minimum on $(0,1)$ at $t = \sqrt{2}-1$. (See Figure 5.8b.) Since g is symmetric in $x_1, x_2, \ldots, x_n$, without loss of generality assume that $x_1 \geq x_2 \geq \ldots \geq x_n$. Proving that g has a unique critical point amounts to showing that if $(x_1, x_2, \ldots, x_n)$ satisfies (5.12) then $x_i \leq \sqrt{2}-1$ for $i = 1, 2, \ldots, n$. Then since h is injective on $(0, \sqrt{2}-1]$ there is a unique solution to (5.12) in this region, namely $x_1 = x_2 = \cdots = x_n = 1/n$.

Suppose $1 > x_1, \ldots, x_k > \sqrt{2}-1$ and $0 < x_{k+1}, \ldots, x_n \leq \sqrt{2}-1$. Then since h is injective on $(0, \sqrt{2}-1]$ and on $[\sqrt{2}-1, 1)$ it follows that $x_1 = \cdots = x_k$ and $x_{k+1} = \cdots = x_n$. Since $\sum_i x_i = 1$,

$$1 = kx_1 + (n-k)x_n \geq k(\sqrt{2}-1)$$

and so $k \le \lfloor 1/(\sqrt{2}-1) \rfloor = 2$. Thus there are two cases to rule out.

If $k = 1$ then $x_2 = \cdots = x_n = (1 - x_1)/(n-1)$. Furthermore,

$$\frac{1}{x_1} + \frac{2}{1-x_1} = h(x_1) = h(x_n) = \frac{n-1}{1-x_1} + \frac{2(n-1)}{n-2+x_1}.$$

The only solutions to this equation are $x_1 = 1/n$ and $x_1 = n-2$. Since the second solution is not in the region of interest, the only possible solution is $x_1 = 1/n < \sqrt{2} - 1$, which is a contradiction.

If $k = 2$ then $x_1 = x_2$ and $x_3 = \cdots = x_n = (1 - 2x_1)/(n-2)$. Furthermore,

$$\frac{1}{x_1} + \frac{2}{1-x_1} = h(x_1) = h(x_n) = \frac{n-2}{1-2x_1} + \frac{2(n-2)}{n-3+2x_1}.$$

The only solutions to this equation are $x_1 = 1/n$ and

$$x_1 = \frac{n-3 \pm \sqrt{(n-7)^2 - 16}}{4} = \frac{n-3 \pm \sqrt{(n-3)(n-11)}}{4}.$$

Just as in the case $k = 1$, we can discard the solution $x_1 = 1/n$ because $1/n < \sqrt{2} - 1$ for all $n \ge 3$. For $3 < n < 11$ the other two solutions are not real, so certainly do not lie in the interval $[\sqrt{2}-1, 1)$. For $n \ge 11$

$$\frac{n-3+\sqrt{(n-7)^2-16}}{4} \ge \frac{n-3}{4} \ge 1$$

and

$$\frac{n-3-\sqrt{(n-7)^2-16}}{4} \ge \frac{n-3-(n-7)}{4} = 1.$$

So neither of these solutions lie in the interval $[\sqrt{2}-1, 1)$.

Hence the only point in the interior of the standard $(n-1)$-simplex that satisfies (5.12) is $x = (1/n, \ldots, 1/n)$.

Finally we show that approaching the boundary of the $(n-1)$-simplex, $f(x) \to 0$, or, equivalently, $g(x) \to -\infty$. There are two cases to consider. The first is when all of the x_i are bounded away from 1, and (at least) one of the x_i approaches 0. The second is when one of the x_i approaches 1.

Consider the case where x_1, say, approaches 0 and all other x_i satisfy $0 \le x_i \le K < 1$. Then

$$\begin{aligned} g(x) &= \ln(x_1) - 2\ln(1-x_1) + \sum_{i=2}^{n} \log(x_i) - 2\sum_{i=2}^{n} \ln(1-x_i) \\ &\le \ln(x_1) - 2\ln(1-x_1) + \ln((1-K)^{2(n-1)}) \to -\infty \end{aligned}$$

as $x_1 \to 0$.

Consider the case where x_1, say, approaches 1. Then let $x_1 = 1 - \delta$ and observe that the other x_i each satisfy $x_i \le \delta$. Then

$$f(x) = \frac{x_1 \prod_{i=2}^{n} x_i}{(1-x_1)^2 \prod_{i=2}^{n} (1-x_i)^2} \le \frac{(1-\delta)\delta^{n-1}}{\delta^2 (1-\delta)^{2(n-1)}} = \frac{\delta^{n-3}}{(1-\delta)^{2(n-2)}}.$$

Since $n \ge 3$, this can be made arbitrarily small by choosing δ small enough. Hence $f(x) \to 0$ in this case. $\square$

Chapter 6

The Gromov-Thurston proof

In this chapter we outline Gromov and Thurston's proof of Mostow Rigidity that appeared in Thurston's Princeton lecture notes [Thu79, Chapter 6], and in Gromov's survey [Gro81] (with each attributing the results to the other). There are now many other accounts of this approach in the literature, (for example, [Mun80], [GP91, Section 3.12], [BP92, Chapter C], and [Rat94, Section 11.6]) all of them a little different in their approach.

Our aim is to present the main ideas, which are quite beautiful, without getting embroiled in too many technical details.

One aspect of this proof of Mostow Rigidity is that if $n \geq 3$, we can gain useful information about a map $\partial f : \partial\mathbb{H}^n \to \partial\mathbb{H}^n$ by examining how it acts on the vertices of regular ideal simplices. As such, an understanding of the volumes of ideal simplices in $\overline{\mathbb{H}}^n$ is an important ingredient in the proof. We briefly examine this is Section 6.1.

Section 6.2 then introduces a remarkable invariant of orientable manifolds called the Gromov norm. Roughly speaking, this is a measure of the complexity of the 'most efficient triangulation' of the manifold.

While the Gromov norm can be described in terms of singular homology, the definition is rather difficult to work with in this setting. It turns out that defining the Gromov norm in terms of a generalization of singular homology, called 'measure homology', makes it much more user-friendly. Section 6.3 gives a brief introduction to measure homology and Section 6.4 gives a definition of the Gromov norm in terms of measure homology.

Remarkably, for closed orientable hyperbolic manifolds, the Gromov norm is proportional to the volume of the manifold. We give a proof of this in Section 6.6. Finally, in Section 6.7 we sketch a proof of Mostow's Rigidity Theorem using the ideas developed in this chapter.

6.1 Simplices of maximum volume

An *n-simplex* in $\overline{\mathbb{H}}^n$ is the convex hull of $n+1$ points in $\overline{\mathbb{H}}^n$. At the heart of the difference between hyperbolic geometry in two dimensions and in dimensions three and above is the fact that all ideal triangles are congruent, yet there is a wealth of different ideal n-simplices for $n \geq 3$.

Definition 6.1. An n-simplex in $\overline{\mathbb{H}}^n$ is *regular* if it is maximally symmetric.

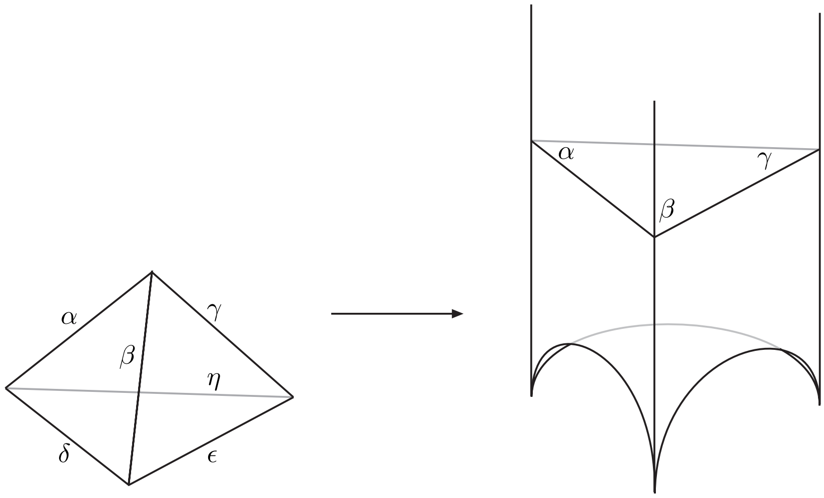


Figure 6.1: Showing that opposite dihedral angles are equal in an ideal tetrahedron.

A key step in the Gromov-Thurston proof of Mostow Rigidity is the following characterisation of simplices of maximal volume.

Theorem 6.2 (Haagerup and Munkholm [HM81])**.** *An n-simplex in $\overline{\mathbb{H}}^n$ is of maximal volume if and only if it is regular and ideal.*

Outline of the proof. Since every finite n-simplex is contained in an ideal n-simplex, only ideal simplices need to be considered. Since all ideal triangles are congruent, there is nothing to left prove in dimension 2.

If $n = 3$, there is a nice formula for the volume of an ideal tetrahedron in terms of its dihedral angles which yields a proof of the theorem. Let $\alpha, \beta, \gamma, \delta, \epsilon, \eta$ be the dihedral angles of an ideal tetrahedron. In the upper half-space model, by applying appropriate isometries, we can arrange for any of the vertices of the tetrahedron to be at infinity (see Figure 6.1). Taking a horospherical cross section in each case, we see that the dihedral angles must satisfy

$$\alpha + \beta + \gamma = \alpha + \delta + \eta = \beta + \delta + \epsilon = \epsilon + \gamma + \eta = \pi.$$

We then deduce that opposite dihedral angles are equal, that is $\alpha = \epsilon$, $\beta = \eta$, and $\gamma = \delta$. The volume of an ideal tetrahedron with dihedral angles α, β, γ then turns out to be exactly

$$L(\alpha) + L(\beta) + L(\gamma)$$

where

$$L(x) = -\int_0^x \log|2 \sin u| \, du$$

is the *Lobachevsky function.* (See the Appendix of [Mil82] for a proof of this fact.) With this established, it is then straightforward to show by Lagrange multipliers that $L(\alpha)+L(\beta)+L(\gamma)$ is maximized (subject to the constraint that $\alpha + \beta + \gamma = \pi$), exactly when $L'(\alpha) = L'(\beta) = L'(\gamma)$. This occurs when $\alpha = \beta = \gamma = \pi/3$, that is when the tetrahedron is regular.

The cases $n > 3$ were first proved in 1981 by Haagerup and Munkholm [HM81]. Their analytic proof, while elementary, is not at all easy and offers little geometric insight. More recently, in 2002 Peyerimhoff [Pey02] used a version of Steiner symmetrization for subsets of $\mathbb{H}^n$ to give another, more geometric, proof. $\square$

6.2 A first attempt at the Gromov norm

Let M be a closed orientable hyperbolic n-manifold. Let $\{C_*(M;\mathbb{R}), \partial\}$ denote the usual singular chain complex of M with coefficients in $\mathbb{R}$. Let $H_*(M;\mathbb{R})$ denote the homology of this chain complex. Given any chain $c \in C_k(M;\mathbb{R})$ where $c = \sum_i c_i\sigma_i$ we define

$$\|c\| = \sum_i |c_i|.$$

The case of most interest to us is when c is a cycle representing a fundamental class $[M] \in H_n(M;\mathbb{R})$ of M. Then we can essentially think of c as a triangulation of M, and $\|c\|$ measures the 'complexity' of the triangulation. In fact if all the coefficients c_i are one or minus one then $\|c\|$ is the number of simplices in the triangulation.

To turn $\|\cdot\|$ into an invariant defined on homology classes define

$$\|[M]\| = \inf\{\|c\| : c \text{ is a cycle representing } [M]\}. \tag{6.1}$$

That is, $\|[M]\|$ tells us something about the lowest complexity of a triangulation of M.

It is fairly clear that to produce a triangulation that is as efficient as possible, we should use simplices that are as large as possible. In dimension 2 we can 'triangulate' closed hyperbolic manifolds with ideal triangles [BP92, Section C.4], so we can essentially find ideal 'triangulations' that achieve the infimum in (6.1). Since it is not generally possible to find ideal 'triangulations' of hyperbolic manifolds by regular simplices in higher dimensions, the Gromov norm, as defined in (6.1) is a difficult invariant to work with.

6.3 Measure homology

In this section we briefly introduce 'measure homology', a homology theory designed by Thurston to make working with the Gromov norm much less cumbersome. The motivation for measure homology is to (vastly) expand the set of possible chains so that it is easier to describe 'efficient' representatives of $[M]$, and hence easier to work with the Gromov norm.

To define measure homology carefully would require wading through quite a bit of technical detail. We will skip over such details, referring the interested reader to Ratcliffe's account [Rat94].

6.3.1 Measure chains

Let $\Delta^n \subset \mathbb{R}^{n+1}$ denote the standard n-simplex and let $\mathcal{S}^n_M$ be the set of smooth singular simplices in M. Equip these with the C^1 topology in which two simplices are 'close' if both the maps and their derivatives are 'close' in the topology of uniform convergence.

Let $\mathcal{C}_k(M)$ be the set of signed Borel measures on $\mathcal{S}^k_M$. To avoid pathological situations, we will also require that the measures have compact support and finite total variation. This set of measures will form the *k-chains* in measure homology.

We can think of these as a generalization of singular k-chains as follows. Suppose $c = \sum_i c_i\sigma_i$ is a smooth singular k-chain. Then the corresponding k-chain in measure homology is $c = \sum_i c_i\delta_{\sigma_i}$ where δ_{σ_i} denotes the measure with a point mass at the smooth singular simplex σ_i.

6.3.2 The boundary operator

Now we define a boundary operator on measure chains. Given $\mu \in \mathcal{C}_{k+1}$, define $\mu^{(i)} \in \mathcal{C}_k$ for $i = 0, 1, \ldots, k+1$ as follows. If S is a Borel subset of $\mathcal{S}^k_M$ then

$$\mu^{(i)}(S) = \mu(\{\text{smooth } (k+1)\text{-simplices whose } i\text{th face is in } S\}).$$

With this definition established the boundary operator $\partial : \mathcal{C}_{k+1} \to \mathcal{C}_k$ in measure homology is defined by

$$\partial(\mu) = \sum_{i=0}^{k+1} (-1)^i \mu^{(i)}.$$

It turns out that $\{\mathcal{C}_*(M), \partial\}$ defines a chain complex. We call the associated homology theory *measure homology*. In particular we denote the nth measure homology vector space by $\mathcal{H}_n(M)$ and if $z \in \mathcal{C}_n(M)$ is a cycle then $[z] \in \mathcal{H}_n(M)$ denotes the homology class of z.

6.3.3 Induced maps on measure homology

Any smooth map $f : M \to N$ between manifolds induces a map $f_\# : \mathcal{S}^n_M \to \mathcal{S}^n_N$ given by $f_\#(\sigma) = f \circ \sigma$. This, in turn, induces a map $f_{\#*} : \mathcal{C}_n(M) \to \mathcal{C}_n(N)$ defined by $\mu \mapsto f_{\#*}[\mu]$, the pushforward of μ by $f_\#$.

As usual, if $f_{\#*}$ is a chain map (i.e. $f_{\#*}\partial = \partial f_{\#*}$) then $f_{\#*}$ descends to a map f_* on measure homology.

6.3.4 Fundamental classes

Remarkably, it turns out that on spaces where they are both defined, measure homology and singular homology coincide [Löh06]. So just as in singular homology, if M is an orientable n-manifold then $\mathcal{H}_n(M)$ is generated by the inclusion of $[M] \in H_n(M; \mathbb{Z})$ into $\mathcal{H}_n(M)$, which we continue to call a fundamental class of M.

Recall that if $z = \sum_i z_i \sigma_i \in C_n(M; \mathbb{R})$ is a singular n-cycle and Ω is the volume form on M we can find out which multiple of $[M]$ that z represents by integration. More precisely,

$$\mathrm{Vol}(M)[z] = \langle z, \Omega \rangle [M] \tag{6.2}$$

where

$$\langle z, \Omega \rangle = \int_z \Omega := \sum_i z_i \int_{\Delta^n} \sigma_i^* \Omega = \sum_i z_i \mathrm{AlgVol}(\sigma_i). \tag{6.3}$$

Here $\mathrm{AlgVol}(\sigma) := \int_{\Delta^n} \sigma^* \Omega$. (Note that (6.2) is well defined because the pairing depends only on the homology class of z by Stokes' theorem.)

An analogous procedure works if $\mu \in \mathcal{C}_n(M)$ is a measure n-cycle. In this case we define the pairing by

$$\langle \mu, \Omega \rangle = \int_\mu \Omega := \int_{\mathcal{S}^n_M} \int_{\Delta^n} \sigma^* \Omega \, d\mu(\sigma) = \int_{\mathcal{S}^n_M} \mathrm{AlgVol}(\sigma) \, d\mu(\sigma). \tag{6.4}$$

Notice that if we think of smooth singular n-chains as weighted combinations of point masses then (6.4) reduces to (6.3).

6.4 The Gromov norm revisited

Now we are in a position to re-define the Gromov norm of a closed orientable manifold, making use of the technology we introduced in the previous section.

Definition 6.3. Let M be a closed orientable n-manifold and let $[M] \in \mathcal{H}_n(M)$ be a fundamental class of M. Then the *Gromov norm* of M is

$$\|[M]\| = \inf\{\|\mu\| : \mu \in \mathcal{C}_n(M) \text{ represents } [M]\}.$$

Recall that by $\|\mu\|$ we mean the total variation of the signed measure μ.

6.5 Straightening and smearing chains

In this section we introduce two constructions that play key roles in what follows.

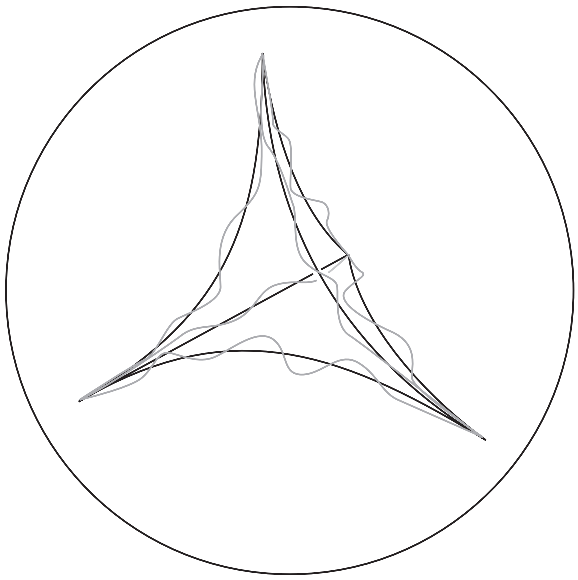


Figure 6.2: Straightening a singular simplex.

6.5.1 Straightening chains

A *straight k-simplex* in $\mathbb{H}^n$ is a smooth singular simplex whose image is the convex hull of $k+1$ points in $\mathbb{H}^n$. The straight k-simplex with vertices $v_0, v_1, \ldots, v_k \in \mathbb{H}^n$ can be explicitly described in a number of ways.

We will give a novel description of straightening using the visual measure and barycentre constructions from Chapter 5. As such, our discussion will be more detailed in this section than elsewhere in this chapter.

For $i = 0, 1, \ldots, k$ let μ_{v_i} be the visual measure on $\partial\mathbb{H}^n$ corresponding to each $v_i \in \mathbb{H}^n$. Since $\mathcal{P}(\partial\mathbb{H}^n)$ is a vector space, we can then define an 'affine simplex' $\Delta^k \to \mathcal{P}(\partial\mathbb{H}^n)$ by

$$(t_0, t_1, \ldots, t_k) \mapsto \sum_{i=0}^{k} t_i \mu_{v_i}$$

(where the t_i are barycentric coordinates for Δ^k). Using the barycentre construction, we can map this back to $\mathbb{H}^n$ to obtain a new k-simplex in $\mathbb{H}^n$. So if σ is a smooth singular k-simplex with vertices at $v_0, v_1, \ldots, v_k \in \mathbb{H}^n$ define $\widetilde{\mathrm{STR}} : \mathcal{S}^k_{\mathbb{H}^n} \to \mathcal{S}^k_{\mathbb{H}^n}$ by

$$\widetilde{\mathrm{STR}}(\sigma)(t_0, t_1, \ldots, t_k) = \mathrm{bar}\left(\sum_{i=0}^{k} t_i \mu_{v_i}\right).$$

Lemma 6.4. *For any $\varphi \in$ Isom$(\mathbb{H}^n)$,*

$$\varphi \circ \widetilde{\mathrm{STR}} = \widetilde{\mathrm{STR}} \circ \varphi.$$

Furthermore, for any smooth k-simplex σ in $\mathbb{H}^n$, $\widetilde{\mathrm{STR}}(\sigma)$ is a straight simplex.

Proof. Let σ be a k-simplex in $\mathbb{H}^n$ with vertices at $v_0, v_1, \ldots, v_k$. Let $\varphi \in \mathrm{Isom}(\mathbb{H}^n)$ be any isometry and let $t_0, t_1, \ldots, t_k > 0$ be such that $\sum_{i=0}^{k} t_i = 1$. Then since $\varphi_*[\mu_x] = \mu_{\varphi(x)}$ for all $x \in \mathbb{H}^n$,

$$\varphi_*[\sum_{i=0}^{m} t_i \mu_{v_i}] = \sum_{i=0}^{m} t_i \varphi_*[\mu_{v_i}] = \sum_{i=0}^{m} t_i \mu_{\varphi(v_i)}.$$

Since $\varphi(\mathrm{bar}\,(\nu)) = \mathrm{bar}\,(\varphi_*[\nu])$ for all atomless measures ν in $\mathcal{P}(\partial\mathbb{H}^n)$,

$$\begin{aligned}
(\varphi \circ \widetilde{\mathrm{STR}})(\sigma)(t_0, t_1, \ldots, t_k) &= \varphi\left[\mathrm{bar}\left(\sum_{i=0}^{m} t_i \mu_{v_i}\right)\right] \\
&= \mathrm{bar}\left(\varphi_*\left[\sum_{i=0}^{m} t_i \mu_{v_i}\right]\right) \\
&= \mathrm{bar}\left(\sum_{i=0}^{m} t_i \mu_{\varphi(v_i)}\right) \\
&= (\widetilde{\mathrm{STR}} \circ \varphi)(\sigma)(t_0, t_1, \ldots, t_k).
\end{aligned}$$

Now we will show that $\widetilde{\mathrm{STR}}(\sigma)$ is actually a straight simplex.

Observe that τ is a straight simplex if and only if all of its faces are straight simplices and every isometry of $\mathbb{H}^n$ that fixes the vertices of τ fixes τ itself. Also note that if $\tau^{(i)}$ denotes the ith face map of τ then

$$\widetilde{\mathrm{STR}}(\tau^{(i)}) = \widetilde{\mathrm{STR}}(\tau)^{(i)}. \tag{6.5}$$

It is clear that any 0-simplex is straight. Arguing by induction, take any k-simplex σ and assume that $\widetilde{\mathrm{STR}}(\sigma)$ is straight. Then if τ is a $k+1$ simplex, its faces $\tau^{(i)}$ are k-simplices. Hence for each i, $\widetilde{\mathrm{STR}}(\tau^{(i)}) = \widetilde{\mathrm{STR}}(\tau)^{(i)}$ is straight, so the faces of $\widetilde{\mathrm{STR}}(\tau)$ are all straight. Now let φ be an isometry that fixes the vertices of τ. Then since straightening only depends on the position of the vertices of the original simplex,

$$\varphi(\widetilde{\mathrm{STR}}(\tau)) = \widetilde{\mathrm{STR}}(\varphi(\tau)) = \widetilde{\mathrm{STR}}(\tau)$$

so $\widetilde{\mathrm{STR}}(\tau)$ is fixed by φ and so is actually straight. □

Now if M is a closed orientable hyperbolic n-manifold we want to define a straightening operation

$$\text{STR} : \mathcal{S}^k_M \to \mathcal{S}^k_M.$$

Since Δ^k is simply connected we can lift any k-simplex σ in M to a k-simplex $\tilde{\sigma}$ in $\mathbb{H}^n$. Then $\widetilde{\text{STR}}(\tilde{\sigma})$ is a straight simplex in $\mathbb{H}^n$. Then define $\text{STR}(\sigma)$ to be the projection of $\widetilde{\text{STR}}(\tilde{\sigma})$ back to M. Since $\widetilde{\text{STR}}$ is equivariant with respect to isometries, this construction is independent of the choice of lift of σ, so $\text{STR}(\cdot)$ is well defined. Furthermore, we can extend STR to a map on $\mathcal{C}_k(M)$ by $\mu \mapsto \text{STR}_*[\mu]$.

Lemma 6.5. *$\widetilde{\text{STR}}$ is a chain map that is homotopic to the identity map on $\mathcal{S}^k_{\mathbb{H}^n}$. Hence* STR *is a chain map that is homotopic to the identity map on $\mathcal{S}^k_M$.*

Proof. The fact that $\widetilde{\text{STR}}$ is a chain map (i.e. $\partial\,\widetilde{\text{STR}} = \widetilde{\text{STR}}\,\partial$) follows immediately from (6.5) and the definition of the boundary operator.

To see that $\widetilde{\text{STR}}$ is homotopic to the identity, let $\tilde{F} : \mathcal{S}^k_{\mathbb{H}^n} \times [0,1] \to \mathcal{S}^k_{\mathbb{H}^n}$ be defined by

$$\tilde{F}(\tilde{\sigma}, s)(t_0, t_1, \ldots, t_k) = \text{bar}\left((1-s)\mu_{\tilde{\sigma}(t_0,t_1,\ldots,t_k)} + s\sum_{i=0}^{k} t_i \mu_{v_i}\right)$$

for any k-simplex $\tilde{\sigma}$ in $\mathbb{H}^n$ with vertices at $v_0, v_1, \ldots, v_k$ and any point $(t_0, t_1, \ldots, t_k)$ in the standard k-simplex (described in barycentric coordinates).

Clearly $\tilde{F}$ is a homotopy and $\tilde{F}(\tilde{\sigma}, 0) = \tilde{\sigma}$ and $\tilde{F}(\tilde{\sigma}, 1) = \widetilde{\text{STR}}(\tilde{\sigma})$.

Since $\widetilde{\text{STR}}$ is equivariant with respect to isometries and the covering projection $\mathbb{H}^n \to M$ is a chain map, it follows that STR is a chain map. Furthermore, $\tilde{F}$ projects to a homotopy $F : \mathcal{S}^k_M \times [0,1] \to \mathcal{S}^k_M$ from the identity on $\mathcal{S}^k_M$ to STR, as claimed. $\square$

Since STR is a chain map, it descends to a map on homology. Moreover, since STR is homotopic to the identity it follows that it induces the identity on homology. Hence any homology class can be represented by a 'straight' cycle, that is a measure supported only on straight simplices.

6.5.2 Smearing chains

Throughout this section, let $M = \mathbb{H}^n/\Gamma$ be a hyperbolic manifold.

Haar measure. Let $G = \text{Isom}_+(\mathbb{H}^n)$. There is a unique positive measure $\tilde{h}$ on G such that for all Borel $H \subset G$ and all $g \in G$,

$$\tilde{h}(gH) = \tilde{h}(H) = \tilde{h}(Hg) \quad \text{and} \quad \tilde{h}(H) = \text{Vol}(H \cdot x)$$

where $H \cdot x$ is the set of images of $x \in \mathbb{H}^n$ under the isometries in H. This measure is called the *Haar measure* on G. A left-invariant Haar measure exists for any locally compact Lie group. In this case the Haar measure is also right-invariant (see [BP92, Section C.4] for a discussion of Haar measure in this context). The bi-invariance of $\tilde{h}$ means that it descends to a measure h on G/Γ which we also refer to by the name Haar measure. The normalization of $\tilde{h}$ is such that $h(G/\Gamma) = \text{Vol}(\mathbb{H}^n/\Gamma)$.

The smearing construction. Our aim is to start with a singular simplex $\sigma \in \mathcal{S}^n_{\mathbb{H}^n}$ and construct a measure chain $\textsc{smr}_M(\sigma) \in \mathcal{C}_n(M)$ that is, essentially, a measure uniformly supported on all projections of isometric copies of σ in M. To do this, let $p : \mathbb{H}^n \to M$ be the covering projection and define a function $\alpha(\sigma) : G/\Gamma \to \mathcal{S}^n_M$ by

$$\alpha(\sigma)(g\Gamma) = p \circ g \circ \sigma.$$

Then define

$$\textsc{smr}_M(\sigma) = \alpha(\sigma)_*[h],$$

the pushforward of the Haar measure on G/Γ by $\alpha(\sigma)$.

Proposition 6.6. *If σ is any smooth simplex in $\mathbb{H}^n$ then $\textsc{smr}_M(\sigma)$ has the following properties:*

1. $\|\textsc{smr}_M(\sigma)\| = \mathit{Vol}(M)$
2. $\langle \textsc{smr}_M(\sigma), \Omega_M \rangle = \mathit{AlgVol}(\sigma)\, \mathit{Vol}(M)$.

where Ω_M is the volume form on M.

Proof. The first assertion follows from the fact that $\textsc{smr}_M(\sigma)$ is a positive measure and satisfies

$$\textsc{smr}_M(\mathcal{S}^n_M) = \alpha(\sigma)_*[h](\mathcal{S}^n_M) = h(\alpha(\sigma)^{-1}[\mathcal{S}^n_M]) = h(G/\Gamma) = \mathrm{Vol}(M).$$

To see that the second property holds, let Ω and Ω_M be the volume forms on $\mathbb{H}^n$ and M respectively. Then using the facts that $p^*\Omega_M = \Omega$ and $g^*\Omega = \Omega$ for any $g \in G$,

$$\begin{aligned}
\langle \textsc{smr}_M(\sigma), \Omega_M \rangle &= \int_{\mathcal{S}^n_M} \left(\int_{\Delta^n} \tau^*\Omega_M \right) d\,\textsc{smr}_M(\sigma)(\tau) \\
&= \int_{G/\Gamma} \left(\int_{\Delta^n} [\alpha(\sigma)(g\Gamma)]^*\Omega_M \right) dh(g\Gamma) \\
&= \int_{G/\Gamma} \left(\int_{\Delta^n} \sigma^* g^* p^* \Omega_M \right) dh(g\Gamma) \\
&= \int_{G/\Gamma} \left(\int_{\Delta^n} \sigma^*\Omega \right) dh(g(\Gamma)) \\
&= \mathrm{AlgVol}(\sigma)\mathrm{Vol}(M).
\end{aligned}$$

□

6.6 Gromov norm and volume are proportional

It is a remarkable theorem of Gromov that the Gromov norm and the volume of a closed orientable hyperbolic n-manifold are proportional to each other. Since the Gromov norm is a homology (and hence homotopy) invariant, it follows that hyperbolic manifolds that are homotopy equivalent must have the same volume. Of course this is a corollary of Mostow Rigidity, and we also came across this corollary in the course of our proof of Mostow Rigidity according to Besson, Courtois, and Gallot. It is then probably little surprise that this result is a key step in the Gromov-Thurston proof of Mostow Rigidity.

The proof we give here relies on the straightening and smearing constructions that we sketched in the previous section. In fact, those constructions were largely motivated by their great utility in streamlining the proof of this result.

Theorem 6.7. *If M is a closed, orientable, hyperbolic n-manifold then*

$$\|[M]\| = \frac{Vol(M)}{v_n}$$

where v_n is the maximum volume of a (straight) n-simplex in $\mathbb{H}^n$.

Proof. First we show that $\|[M]\| \geq \mathrm{Vol}(M)/v_n$. Let Ω_M be the volume form for M. Let $\mu \in \mathcal{C}_n(M)$ be a cycle that represents a fundamental class $[M]$ of M. Then $\mathrm{STR}_*(\mu)$ represents the same fundamental class of M. Since projection of simplices from $\mathbb{H}^n$ to M never increases volume, any straight simplex in M has volume at most v_n. Thus

$$\begin{aligned}
\mathrm{Vol}(M) = \int_M \Omega_M &= \int_{\mathcal{S}_M^n} \left(\int_{\Delta^n} \sigma^* \Omega_M \right) d\,\mathrm{STR}_*(\mu)(\sigma) \\
&\leq v_n \int_{\mathcal{S}_M^n} d\,\mathrm{STR}_*(\mu)(\sigma) \\
&\leq v_n \|\mathrm{STR}_*(\mu)\| \\
&\leq v_n \|\mu\|.
\end{aligned}$$

Taking the infimum over all μ representing $[M]$ gives the result.

Now for the reverse inequality. Given any positively oriented straight n-simplex σ in $\mathbb{H}^n$, let σ_- denote the image of σ under some reflection. Let

$$\mu = \frac{1}{2} \left(\mathrm{SMR}_M(\sigma) - \mathrm{SMR}_M(\sigma_-) \right) \in \mathcal{C}_n(M).$$

Then since $\mathrm{SMR}_M(\sigma)$ and $\mathrm{SMR}_M(\sigma_-)$ have disjoint support,

$$\|\mu\| = \frac{1}{2} \|\mathrm{SMR}_M(\sigma)\| + \frac{1}{2} \|\mathrm{SMR}_M(\sigma_-)\| = \mathrm{Vol}(M)$$

from the first part of Proposition 6.6.

Next we argue that μ is actually a cycle. Let $p : \mathbb{H}^n \to M$ denote the covering projection. Then for every face of every isometric copy of $p \circ \sigma$ (on which $\mathrm{SMR}_M(\sigma)$ is supported), there is a face of an isometric copy of $p \circ \sigma_-$ (on which $\mathrm{SMR}_M(\sigma_-)$ is supported) that matches the first face, but with opposite orientation. Hence the faces cancel out in pairs and so $\partial(\mu) = 0$.

Applying the second part of Proposition 6.6 gives

$$\begin{aligned}
[\mu] &= \frac{1}{\mathrm{Vol}(M)} \langle \mu, \Omega_M \rangle [M] \\
&= \frac{1}{2\mathrm{Vol}(M)} \left(\langle \mathrm{SMR}_M(\sigma), \Omega_M \rangle - \langle \mathrm{SMR}_M(\sigma_-), \Omega_M \rangle \right) [M] \\
&= \frac{1}{2\mathrm{Vol}(M)} \left(\mathrm{AlgVol}(\sigma)\mathrm{Vol}(M) - \mathrm{AlgVol}(\sigma_-)\mathrm{Vol}(M) \right) [M] \\
&= \mathrm{AlgVol}(\sigma)[M].
\end{aligned}$$

Note that for straight n-simplices σ, $\mathrm{AlgVol}(\sigma) = \pm\mathrm{Vol}(\sigma(\Delta^n))$ where the sign is chosen according to whether σ is positively or negatively oriented. So for any straight simplex σ in $\mathbb{H}^n$, $\mu/\mathrm{Vol}(\sigma)$ represents $[M]$ and so

$$\|[M]\| \leq \frac{\|\mu\|}{\mathrm{Vol}(\sigma)} = \frac{\mathrm{Vol}(M)}{\mathrm{Vol}(\sigma)}.$$

Taking the infimum over straight simplices σ in $\mathbb{H}^n$ gives

$$\|[M]\| \leq \frac{\mathrm{Vol}(M)}{v_n}. \qquad \square$$

6.7 A proof of Mostow's theorem

Given a homotopy equivalence f between closed hyperbolic manifolds M_1 and M_2, we can construct a map $\bar{f} : \overline{\mathbb{H}}^n \to \overline{\mathbb{H}}^n$ which is smooth on $\mathbb{H}^n$, whose restriction to $\partial\mathbb{H}^n$ is a homeomorphism, and which is equivariant with respect to the action of $\pi_1(M_1)$ and $\pi_1(M_2)$ on $\mathbb{H}^n$.

Such a map exists by a similar construction to the one detailed in Chapter 4. The idea is to first replace the homotopy equivalence f with a smooth homotopy equivalence in the same homotopy class (see [Lee03, Theorem 10.21] for a proof). Then lift f to a smooth map $\tilde{f} : \mathbb{H}^n \to \mathbb{H}^n$ with the appropriate equivariance property. Using the compactness of M_1 and M_2 it can be shown that $\tilde{f}$ is a quasi-isometry and so, using arguments from Chapter 4, extends to a continuous map $\bar{f} : \overline{\mathbb{H}}^n \to \overline{\mathbb{H}}^n$ whose restriction $\partial\tilde{f}$ to $\partial\mathbb{H}^n$ is a homeomorphism.

As in the proof of Besson, Courtois, and Gallot, it remains to show that $\partial\tilde{f}$ is actually the restriction of an isometry to $\partial\mathbb{H}^n$. To achieve this, first we establish that $\partial\tilde{f}$ takes the vertices of maximal volume simplices in $\overline{\mathbb{H}}^n$ to the vertices of maximal volume simplices in $\overline{\mathbb{H}}^n$. Second, we use the characterization of simplices of maximal volume given in Section 6.1 to establish that $\partial\tilde{f}$ takes the vertices of regular ideal simplices to the vertices of regular ideal simplices. Finally, if $n \geq 3$, we show that this implies $\partial\tilde{f}$ is the restriction of an isometry to $\partial\mathbb{H}^n$.

We now give a more detailed account of the first and third of the steps in this strategy.

Proposition 6.8. *If $u_0, u_1, \ldots, u_n \in \partial\mathbb{H}^n$ are the vertices of a simplex of maximal volume in $\overline{\mathbb{H}}^n$ then $\partial\tilde{f}(u_0), \partial\tilde{f}(u_1), \ldots, \partial\tilde{f}(u_n)$ are the vertices of a simplex of maximal volume in $\overline{\mathbb{H}}^n$.*

Proof. (Our proof follows that given by Munkholm in [Mun80].) We argue by contradiction. Suppose that $u_0, u_1, \ldots, u_n \in \partial\mathbb{H}^n$ are the vertices of an ideal simplex τ of maximal volume but $\partial\tilde{f}(u_0), \partial\tilde{f}(u_1), \ldots, \partial\tilde{f}(u_n)$ are not. Then, by continuity, simplices that are sufficiently close to τ have images under $\tilde{f}$ which, after straightening, have volumes that are bounded away from v_n.

More precisely, take $\epsilon > 0$ and let $U_0, U_1, \ldots, U_n \subset \mathbb{H}^n$ be neighbourhoods of the u_i so that if σ is a (necessarily finite) simplex with ith vertex in U_i then

$$\operatorname{Vol}(\widetilde{\text{STR}}(\tilde{f}(\sigma))) \leq v_n - \epsilon \quad \text{for some} \quad \epsilon > 0. \tag{6.6}$$

For convenience, let us call simplices with one vertex in each U_i that satisfy (6.6) 'bad' simplices. Let

$$B = \{g\Gamma : g \text{ takes 'bad' simplices to 'bad' simplices}\} \subset G/\Gamma.$$

It turns out that the Haar measure of B is positive. In symbols $h(B) =: m > 0$ where h denotes the Haar measure on G/Γ.

Now choose a positively oriented 'bad' straight simplex σ in $\mathbb{H}^n$ with volume $\operatorname{Vol}(\sigma) > v_n - \delta$ for some δ (depending only on ϵ and m) to be chosen later. We can do this because, rather oxymoronically, we started with the assumption that a 'bad' ideal simplex exists!

If $g\Gamma \in B$ then since σ is 'bad', so is $g \circ \sigma$. Hence

$$\operatorname{Vol}(\widetilde{\text{STR}}(\tilde{f}(g \circ \sigma))) \leq v_n - \epsilon \leq \operatorname{Vol}(\sigma) + \delta - \epsilon. \tag{6.7}$$

Similarly if $g\Gamma \notin B$ then

$$\operatorname{Vol}(\widetilde{\text{STR}}(\tilde{f}(g \circ \sigma))) \leq v_n \leq \operatorname{Vol}(\sigma) + \delta. \tag{6.8}$$

As in the proof of Theorem 6.7, let $\mu = \frac{1}{2}[\text{SMR}_M(\sigma) - \text{SMR}_M(\sigma_-)]$ where σ_- is a reflected copy of σ. We can produce a cycle in $\mathcal{C}_n(N)$ by pushing forward μ by $\text{STR} \circ f_\#$, where $f : M \to N$ is the homotopy equivalence between M and N that lifts to $\tilde{f}$.

We will now find out which multiple of $[N]$ is represented by $\textsc{str}_* f_{\#*}[\mu]$. Now if $p : \mathbb{H}^n \to M$ denotes the covering projection,

$$\begin{aligned}
\langle \textsc{str}_* f_{\#*}[\textsc{smr}_M(\sigma)], \Omega_N\rangle &= \int_{\mathcal{S}^n_N} \left(\int_{\Delta^n} \tau^*\Omega_n\right) \, d\,\textsc{str}_* f_{\#*}\alpha(\sigma)_*[h](\tau) \\
&= \int_{G/\Gamma} \left(\int_{\Delta^n} (\textsc{str} \circ f \circ p \circ g \circ \sigma)^*\Omega_N\right) \, d\,h(g\Gamma) \\
&= \int_{G/\Gamma} \left(\int_{\Delta^n} (\textsc{str} \circ p \circ \tilde{f} \circ g \circ \sigma)^*\Omega_N\right) \, d\,h(g\Gamma) \\
&= \int_{G/\Gamma} \left(\int_{\Delta^n} (p \circ \widetilde{\textsc{STR}} \circ \tilde{f} \circ g \circ \sigma)^*\Omega_N\right) \, d\,h(g\Gamma) \\
&= \int_{G/\Gamma} \left(\int_{\Delta^n} (\widetilde{\textsc{STR}} \circ \tilde{f} \circ g \circ \sigma)^*\Omega\right) \, d\,h(g\Gamma) \\
&= \int_{B} \left(\int_{\Delta^n} (\widetilde{\textsc{STR}} \circ \tilde{f} \circ g \circ \sigma)^*\Omega\right) \, d\,h(g\Gamma) + \\
&\qquad \int_{(G/\Gamma)\setminus B} \left(\int_{\Delta^n} (\widetilde{\textsc{STR}} \circ \tilde{f} \circ g \circ \sigma)^*\Omega\right) \, d\,h(g\Gamma) \\
&\leq m(\mathrm{AlgVol}(\sigma) + \delta - \epsilon) + (\mathrm{Vol}(M) - m)(\mathrm{AlgVol}(\sigma) + \delta) \\
&= \mathrm{Vol}(M)(\mathrm{Vol}(\sigma) + \delta) - m\epsilon
\end{aligned}$$

where we have used (6.7) and (6.8) and the fact that σ is positively oriented (so $\mathrm{AlgVol}(\sigma) = \mathrm{Vol}(\sigma) \geq 0$) to produce the final estimate. Choosing $\delta < \epsilon m/\mathrm{Vol}(M)$ gives

$$\langle \textsc{str}_* f_{\#*}[\textsc{smr}_M(\sigma)], \Omega_N\rangle < \mathrm{Vol}(\sigma)\mathrm{Vol}(M).$$

Similarly, since σ_- is negatively oriented, a similar calculation shows that

$$-\langle \textsc{str}_* f_{\#*}[\textsc{smr}_M(\sigma_-)], \Omega_N\rangle < -\mathrm{AlgVol}(\sigma_-)\mathrm{Vol}(M) = \mathrm{Vol}(\sigma)\mathrm{Vol}(M).$$

Hence

$$\langle \textsc{str}_* f_{\#*}[\mu], \Omega_N\rangle < \mathrm{Vol}(\sigma)\mathrm{Vol}(M). \tag{6.9}$$

During the course of the proof of Theorem 6.7, we established that $[\mu] = \mathrm{Vol}(\sigma)[M]$. Since Theorem 6.7 implies that homotopy equivalent closed hyperbolic manifolds have the same volume, $\mathrm{Vol}(M) = \mathrm{Vol}(N)$. So (6.9) implies that $\textsc{str}_* f_{\#*}[\mu] = k[N]$ where

$$|k| < \mathrm{Vol}(\sigma)\frac{\mathrm{Vol}(M)}{\mathrm{Vol}(N)} = \mathrm{Vol}(\sigma).$$

Since $\textsc{str} \circ f_\#$ is a homotopy equivalence, it is a map of degree ± 1. Then we have the following contradiction:

$$k[N] = \textsc{str}_* f_{\#*}[\mu] = \mathrm{Vol}(\sigma)\textsc{str}_* f_{\#*}[M] = \pm\mathrm{Vol}(\sigma)[N].$$

□

Before completing the proof, we need a little lemma about regular ideal simplices. Note that this is the only point where we use the assumption that $n \geq 3$ in this proof.

Lemma 6.9. *If $n \geq 3$ and $v_0, v_1, \ldots v_{n-1}, v_n$ are the vertices of a regular ideal n-simplex σ then the only other regular ideal simplex with vertices at $v_0, v_1, \ldots, v_{n-1}$ is the reflection of σ in the hyperplane containing $v_0, v_1, \ldots, v_{n-1}$.*

Proof. Consider the half-space model and note that isometries of $\mathbb{H}^n$ act on the boundary by Euclidean similarities. Assume, without loss of generality, that $v_0 = \infty$. Denote by $E(\sigma)$ the Euclidean simplex spanned by the n vertices of σ that are not at infinity. Since σ is regular (and so is $\varphi(\sigma)$ for any isometry φ), it follows that $E(\sigma)$ and $E(\varphi(\sigma))$ are regular Euclidean simplices and are related by a similarity. If σ and $\varphi(\sigma)$ share a (vertical) face then $E(\sigma)$ and $E(\varphi(\sigma))$ share a face. Since $n \geq 3$ this means that $E(\sigma)$ and $E(\varphi(\sigma))$ must be related by a Euclidean isometry. Since there are exactly two such Euclidean isometries that fix a face of $E(\sigma)$ (the identity and the reflection in that face), the result follows. $\square$

Now we present the final step in proving Mostow's Theorem using the methods of Gromov and Thurston.

Proposition 6.10. *If $n \geq 3$ and $\partial\tilde{f} : \partial\mathbb{H}^n \to \partial\mathbb{H}^n$ takes vertices of regular ideal simplices to vertices of regular ideal simplices then $\partial\tilde{f} = \varphi|_{\partial\mathbb{H}^n}$ for some $\varphi \in Isom(\mathbb{H}^n)$.*

Proof. Choose any geodesic ray $\beta : [0, \infty) \to \mathbb{H}^n$ with $\beta(0) = O$.

Choose any regular ideal simplex σ_1 containing O and let its vertices be $v_0, v_1, \dots, v_n$. Let $u_i = \partial\tilde{f}(v_i)$ for $i = 0, 1, \dots, n$. Since the u_i also span a regular ideal simplex, there is an isometry $\psi \in \mathrm{Isom}(\mathbb{H}^n)$ such that $\psi \circ \partial\tilde{f}$ fixes each v_i. It remains to show that $\psi \circ \partial\tilde{f}$ fixes $\partial\mathbb{H}^n$.

Let H_1 denote the face of σ_1 that β intersects. Then define φ_1 to be the reflection in H_1 and let $\sigma_2 = \varphi_1(\sigma_1)$. Then since $\psi \circ \partial\tilde{f}$ fixes the vertices of the common face of σ_1 and σ_2, and $\psi \circ \partial\tilde{f}$ sends vertices of regular ideal simplices to vertices of regular ideal simplices, it follows that $\psi \circ \partial\tilde{f}$ also fixes the vertices of σ_2.

Let $H_2 \neq H_1$ denote another face (possibly of dimension less than $n-1$) of σ_2 that β intersects. If this face is a vertex of σ_2, then $\psi \circ \partial\tilde{f}$ fixes $\beta(\infty)$ and so we are done. Otherwise, continue the process, constructing a sequence of simplices σ_n whose vertices converge to $\beta(\infty)$. Since $\psi \circ \partial\tilde{f}$ fixes the vertices of each σ_n, it follows by continuity that $\psi \circ \partial\tilde{f}$ fixes $\beta(\infty)$ and so we are done. $\square$

6.8 Thurston's generalization of Mostow Rigidity

In his Princeton lecture notes [Thu79, Chapter 6] Thurston proves a significant generalization of Mostow Rigidity that is similar form to the more general results proved by Besson, Courtois, and Gallot that we outlined in Section 5.6.

Theorem 6.11. *Let $f : M_1 \to M_2$ be any map of non-zero degree between closed oriented hyperbolic n-manifolds ($n \geq 3$) such that*

$$Vol(M_1) = |\deg(f)|\, Vol(M_2).$$

Then f is homotopic to a map which is a local isometry. If $|\deg(f)| = 1$, then f is a homotopy equivalence. Otherwise it is homotopic to a covering map.

Interestingly, the techniques involved in the proof are rather similar to those used in the Gromov-Thurston proof of Mostow's Theorem. The main stumbling block in this more general setting is that under these hypotheses one cannot extend f to a homeomorphism of the boundary of hyperbolic space. Indeed, initially it is only possible to show that the map 'at infinity' is measurable. Nevertheless, Thurston manages to adapt the proof techniques to deal with this situation.

Bibliography

[Aga83] S. Agard, *A geometric proof of Mostow's rigidity theorem for groups of divergence type*, Acta Math. **151** (1983), no. 3-4, 231–252.

[Bau01] H. Bauer, *Measure and integration theory*, de Gruyter Studies in Mathematics, vol. 26, Walter de Gruyter & Co., Berlin, 2001, Translated from the German by Robert B. Burckel.

[BCG95] G. Besson, G. Courtois, and S. Gallot, *Entropies et rigidités des espaces localement symétriques de courbure strictement négative*, Geom. Funct. Anal. **5** (1995), no. 5, 731–799.

[BCG96] ———, *Minimal entropy and Mostow's rigidity theorems*, Ergodic Theory Dynam. Systems **16** (1996), no. 4, 623–649.

[BH99] M.R. Bridson and A. Haefliger, *Metric spaces of non-positive curvature*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 319, Springer-Verlag, Berlin, 1999.

[BP92] R. Benedetti and C. Petronio, *Lectures on hyperbolic geometry*, Universitext, Springer-Verlag, Berlin, 1992.

[Bro60] E.J. Brody, *The topological classification of the lens spaces*, Ann. of Math. (2) **71** (1960), 163–184.

[Cal61] E. Calabi, *On compact, Riemannian manifolds with constant curvature. I*, Proc. Sympos. Pure Math., Vol. III, American Mathematical Society, Providence, R.I., 1961, pp. 155–180.

[CC92] J.W. Cannon and D. Cooper, *A characterization of cocompact hyperbolic and finite-volume hyperbolic groups in dimension three*, Trans. Amer. Math. Soc. **330** (1992), no. 1, 419–431.

[CFKP97] J.W. Cannon, W.J. Floyd, R. Kenyon, and W.R. Parry, *Hyperbolic geometry*, Flavors of geometry, Math. Sci. Res. Inst. Publ., vol. 31, Cambridge Univ. Press, Cambridge, 1997, pp. 59–115.

[DE86] A. Douady and C.J. Earle, *Conformally natural extension of homeomorphisms of the circle*, Acta Math. **157** (1986), no. 1-2, 23–48.

[dR50] G. de Rham, *Complexes à automorphismes et homéomorphie différentiable*, Ann. Inst. Fourier Grenoble **2** (1950), 51–67 (1951).

[Far97] B. Farb, *The quasi-isometry classification of lattices in semisimple Lie groups*, Math. Res. Lett. **4** (1997), no. 5, 705–717.

[GP91] M. Gromov and P. Pansu, *Rigidity of lattices: an introduction*, Geometric topology: recent developments (Montecatini Terme, 1990) (Berlin), Lecture Notes in Math., vol. 1504, Springer, 1991, pp. 39–137.

[Gro81] Michael Gromov, *Hyperbolic manifolds (according to Thurston and Jørgensen)*, Bourbaki Seminar, Vol. 1979/80, Lecture Notes in Math., vol. 842, Springer, Berlin, 1981, pp. 40–53.

[Gro87] M. Gromov, *Hyperbolic groups*, Essays in group theory, Math. Sci. Res. Inst. Publ., vol. 8, Springer, New York, 1987, pp. 75–263.

[Hat02] A. Hatcher, *Algebraic topology*, Cambridge University Press, Cambridge, 2002.

[HM81] U. Haagerup and H.J. Munkholm, *Simplices of maximal volume in hyperbolic n-space*, Acta Math. **147** (1981), no. 1-2, 1–11.

[Iva96] N.V. Ivanov, *Action of Möbius transformations on homeomorphisms: stability and rigidity*, Geom. Funct. Anal. **6** (1996), no. 1, 79–119.

[Lee97] J.M. Lee, *Riemannian manifolds*, Graduate Texts in Mathematics, vol. 176, Springer-Verlag, New York, 1997, An introduction to curvature.

[Lee03] ______, *Introduction to smooth manifolds*, Graduate Texts in Mathematics, vol. 218, Springer-Verlag, New York, 2003.

[Löh06] C. Löh, *Measure homology and singular homology are isometrically isomorphic*, Math. Z. **253** (2006), no. 1, 197–218.

[Mil82] J. Milnor, *Hyperbolic geometry: the first 150 years*, Bull. Amer. Math. Soc. (N.S.) **6** (1982), no. 1, 9–24.

[Mor21] H.M. Morse, *Recurrent geodesics on a surface of negative curvature*, Trans. Amer. Math. Soc. **22** (1921), no. 1, 84–100.

[Mos68] G.D. Mostow, *Quasi-conformal mappings in n-space and the rigidity of hyperbolic space forms*, Inst. Hautes Études Sci. Publ. Math. (1968), no. 34, 53–104.

[Mos73] ______, *Strong rigidity of locally symmetric spaces*, Princeton University Press, Princeton, N.J., 1973, Annals of Mathematics Studies, No. 78.

[MT98] K. Matsuzaki and M. Taniguchi, *Hyperbolic manifolds and Kleinian groups*, Oxford Mathematical Monographs, The Clarendon Press Oxford University Press, New York, 1998, Oxford Science Publications.

[Mun80] H.J. Munkholm, *Simplices of maximal volume in hyperbolic space, Gromov's norm, and Gromov's proof of Mostow's rigidity theorem (following Thurston)*, Topology Symposium, Siegen 1979 (Proc. Sympos., Univ. Siegen, Siegen, 1979) (Berlin), Lecture Notes in Math., vol. 788, Springer, 1980, pp. 109–124.

[Nic91] P.J. Nicholls, *A measure on the limit set of a discrete group*, Ergodic theory, symbolic dynamics, and hyperbolic spaces (Trieste, 1989), Oxford Sci. Publ., Oxford Univ. Press, New York, 1991, pp. 259–297.

[Pan97] P. Pansu, *Volume, courbure et entropie (d'après G. Besson, G. Courtois et S. Gallot)*, Astérisque (1997), no. 245, Exp. No. 823, 3, 83–103, Séminaire Bourbaki, Vol. 1996/97.

[Pey02] N. Peyerimhoff, *Simplices of maximal volume or minimal total edge length in hyperbolic space*, J. London Math. Soc. (2) **66** (2002), no. 3, 753–768.

[Pra73] G. Prasad, *Strong rigidity of* $\mathbf{Q}$*-rank* 1 *lattices*, Invent. Math. **21** (1973), 255–286.

[Rat94] J.G. Ratcliffe, *Foundations of hyperbolic manifolds*, Graduate Texts in Mathematics, vol. 149, Springer-Verlag, New York, 1994.

[Sto06] P.A. Storm, *The minimal entropy conjecture for nonuniform rank one lattices*, Geom. Funct. Anal. **16** (2006), no. 4, 959–980.

[Thu79] W.P. Thurston, *The geometry and topology of three-manifolds*, lecture notes, Princeton University, 1979.

[Thu97] ———, *Three-dimensional geometry and topology. Vol. 1*, Princeton Mathematical Series, vol. 35, Princeton University Press, Princeton, NJ, 1997, Edited by Silvio Levy.

[Tuk85] P. Tukia, *Differentiability and rigidity of Möbius groups*, Invent. Math. **82** (1985), no. 3, 557–578.

[Wei62] A. Weil, *On discrete subgroups of Lie groups. II*, Ann. of Math. (2) **75** (1962), 578–602.

[Whi41] J.H.C. Whitehead, *On incidence matrices, nuclei and homotopy types*, Ann. of Math. (2) **42** (1941), 1197–1239.