Mostow Rigidity
Kleinian Groups Fall 2024

Aaratrick Basu

Abstract

The Mostow Rigidity Theorem states that if M_1 and M_2 are two closed, connected, oriented hyperbolic manifolds of dimension $n \geq 3$ that are homotopy equivalent, then they must be isometric. This is in stark contrast with the behaviour of hyperbolic surfaces, where there are uncountably many non-isometric hyperbolic structures on any surface. These are the notes accompanying my final presentation for a course on Kleinian Groups in Fall 2024, where I presented the proof of Mostow's result due to Gromov, introducing the notion of Gromov norm. As an application, I also proved the result that $\mathrm{Out}(\pi_1(M)) \simeq \mathrm{Isom}(M)$ and is a finite group, for any closed hyperbolic manifold of dimension at least 3. The technical details of the proofs are described in these notes.

Contents

1 Introduction

These notes discuss the details of Gromov's proof of Mostow rigidity, as presented in chapter C of [BP92], with certain details and calculations adapted from [Thu78] and [Mar22].

Theorem 1 (Mostow, 1973)

Let $n \geq 3$ and M_1, M_2 be two $n-$dimensional compact connected oriented hyperbolic manifolds. If $f : M_1 \to M_2$ is a homotopy equivalence, there exists an isometry $q : M_1 \to M_2$ that is homotopic to f.

In fact, we can formulate the result in a shaper form as follows.

Theorem 2

Let $M_i \simeq \mathbb{H}^n/\Gamma_i, i = 1, 2$, be as above. If there is a group isomorphism $\varphi : \Gamma_1 \to \Gamma_2$, then there is an isometry $q \in \text{Isom}(\mathbb{H}^n)$ such that

$$q \circ \gamma = \varphi(\gamma) \circ q,$$

holds for all $\gamma \in \Gamma_1$. In particular, q induces an isometry $\widetilde{\varphi} : M_1 \to M_2$ for which $\widetilde{\varphi}_* = \varphi$.

This version actually follows immediately from Theorem (1) by noting that hyperbolic manifolds are Eilenberg-MacLane spaces because $\mathbb{H}^n$ is contractible, and the fact that any morphism of fundamental groups of such spaces is induced by some continuous map.

1.1 Methods of Proof

There are now several different proofs of Mostow's result and related extensions, each highlighting a different aspect of the rigidity of hyperbolic manifolds in dimension 3 and above. Mostow's original proof involved a lot of analytic techniques and was also quite long and tedious. Modern proofs often start with the following: Consider the lift (to the universal cover $\mathbb{H}^n$) $\widetilde{f}$ of f,

$$\begin{array}{ccc} \mathbb{H}^n & \xrightarrow{\widetilde{f}} & \mathbb{H}^n \\ {\scriptstyle p_1}\downarrow & & \downarrow{\scriptstyle p_2} \\ M_1 & \xrightarrow[f]{} & M_2 \end{array}$$

Note that $\widetilde{f} \circ \gamma = f_*(\gamma) \circ \widetilde{f}$ holds on $\mathbb{H}^n$, for all $\gamma \in \Gamma_1$, for a suitable choice of basepoints. The first step of the proof is then to show that $\widetilde{f}$ can be taken to be a quasi-isometry that extends to a continuous map $\widetilde{f} : \overline{\mathbb{H}^n} \to \overline{\mathbb{H}^n}$, such that it is an injection when restricted to $\partial\mathbb{H}^n$ and the relation $\widetilde{f} \circ \gamma = f_*(\gamma) \circ \widetilde{f}$ holds on all of $\overline{\mathbb{H}^n}$, for each element $\gamma \in \Gamma_1$. We mention here the subsequent steps of three different proofs that start like this:

(i) Gromov's proof: $\widetilde{f}\big|_{\partial\mathbb{H}^n}$ is induced by an isometry, which then descends to an isometry of the manifolds that is homotopic to f. This is shown by looking at images of ideal simplices, and using the Gromov norm. We expose the details of this proof in the sequel.

(ii) Tukia's proof: This follows essentially the same strategy as before, but instead of topological ideas it uses analytic techniques from the theory of quasi-conformal maps. The theorem that is the key tool is:

Theorem 3 (Tukia, 1985)

Suppose Γ_1 is any discrete subgroup of $SO_0(n, 1)$, for which ξ is a conical point. Let $h : \mathbb{S}^{n-1} \to \mathbb{S}^{n-1}$ be a homeomorphism which is differentiable at ξ with non-zero derivative and $h\Gamma_1 h^{-1} \subset SO_0(n, 1)$. Then h is an isometry, i.e a Möbius transformation.

(iii) Besson-Courtois-Gallot's proof: This is probably the fastest proof of Mostow's result right now, and indeed the more general result below is shown.

Theorem 4 (Besson-Courtois-Gallot, 1995)

Let (Y, g) and (X, g_0) be two compact and negatively curved Riemannian manifolds, such that (X, g_0) is hyperbolic. Suppose X and Y are homotopy equivalent, and $\dim X = \dim Y = n \geq 3$. Then,

1) $h(g)^n \operatorname{vol}(Y, g) \geq h(g_0)^n \operatorname{vol}(X, g_0)$

2) Equality holds iff X is isometric to Y, upto scaling of the metric.

Here h denotes the so-called *volume entropy*, defined as

$$h(g) = \lim_{R\to\infty} \frac{1}{R} \log \operatorname{vol}(B(p, R))$$

where $B(p, R)$ is the ball of radius R around any fixed point p in the universal cover. The proof of this result requires quite different techniques than the two above, but still involves investigating the behaviour on the boundary of the relevant induced map.

1.2 Outline of Gromov's Proof

Before diving into the details, we sketch the outline of Gromov's strategy.

(i) Extend $\widetilde{f}$ to the boundary $\partial\mathbb{H}^n$ as mentioned above.

$$\begin{tikzcd}
\mathbb{H}^n \arrow[r, "\widetilde{f}"] \arrow[d, "p_1"'] & \mathbb{H}^n \arrow[d, "p_2"] \\
M_1 \arrow[r, "f"'] & M_2
\end{tikzcd}$$

(ii) Show that the volume function vol() attains its supremum v_n over all geodesic $n-$simplices at the regular and ideal $n-$simplex.

(iii) If $\{u_0, \ldots, u_n\}$ are the vertices of a simplex of volume v_n, then the simplex on $\left\{\widetilde{f}(u_0), \ldots, \widetilde{f}(u_n)\right\}$ also has volume v_n.

(iv) Show that the above fact implies that $\widetilde{f}$ is induced by an isometry of $\mathbb{H}^n$.

Historically, step (ii) was only known for $n \leq 3$ when Gromov published his proof, and is due to Haagerup and Munkholm in the general case. The proof of the result is not very difficult, but still non-trivial. Step (iii) is the most technical of all, and is what requires the notion of Gromov norm of manifolds and Gromov's theorem which relates this quantity to the volume in the case of compact hyperbolic manifolds. Interestingly, step (iv) is the only step where the assumption $n \geq 3$ is needed, and fails for $n = 2$ because there is too little space for the argument to work in some sense.

2 First step of the proof

We start with a homotopy equivalence $f : M_1 \to M_2$. It is a general fact from differential topology that a continuous homotopy can be replaced with a smooth one, and so we assume that f is a smooth homotopy equivalence without loss of generality.

Lemma 1

Let $\widetilde{f} : \mathbb{H}^n \to \mathbb{H}^n$ be the lift of f. Then, $\widetilde{f}$ is a pseudo-isometry.

Here *pseudo-isometry* refers to a slightly more restrictive version of quasi-isometry; F is a $(K, C)-$pseudo-isometry if, for all x, y,

$$\frac{1}{K}d(x,y) - C \leq d(F(x), F(y)) \leq Kd(x,y).$$

Proof. Since M_1 is compact, f has some finite maximum dilatation C. $\widetilde{f}$ is obtained by lifting f, and so behaves locally like f and in particular, must have the same maximum dilatation C. If $g : M_2 \to M_1$ is the homotopy inverse of f, the same also holds for g. Hence, there is some $C_1 > 0$ such that

$$d(\widetilde{f}(x_1), \widetilde{f}(x_2)) \leq C_1 d(x_1, x_2)$$
$$d(\widetilde{g}(x_1), \widetilde{g}(x_2)) \leq C_1 d(x_1, x_2),$$

holds for all $x_1, x_2 \in \mathbb{H}^n$. $\widetilde{g} \circ \widetilde{f}$ is a composition of lifts and hence commutes with the action of Γ_1. The quotient $M_1 = \mathbb{H}^n/\Gamma_1$ is a compact manifold, and so there is a compact Dirichlet domain for Γ_1. Therefore, $\widetilde{g} \circ \widetilde{f}$ has maximum displacement bounded by some $K > 0$, where K is the maximum displacement of points in such a domain. Combining all of this we get

$$d(x_1, x_2) - 2K \leq d(\widetilde{g}(\widetilde{f}(x_1)), \widetilde{g}(\widetilde{f}(x_2))) \leq C_1 d(\widetilde{f}(x_1), \widetilde{f}(x_2)),$$

for all $x_1, x_2 \in \mathbb{H}^n$. This exactly means that $\widetilde{f}$ is a pseudo-isometry with constants $\left(C_1, \frac{2K}{C_1}\right)$. $\square$

2.1 Action of pseudo-isometries at infinity

Recall that we want to prove that $\widetilde{f}$ can be extended to a continuous map $\overline{\mathbb{H}^n} \to \overline{\mathbb{H}^n}$ that is an injection on the boundary. Given that we now know that $\widetilde{f}$ is a pseudo-isometry, this follows from the following more general result.

Theorem 5

Every pseudo-isometry $F : \mathbb{H}^n \to \mathbb{H}^n$ extends continuously to $\overline{\mathbb{H}^n}$, and the extension injects $\partial\mathbb{H}^n$ to itself.

We note here that once we extend $\widetilde{f}$ continuously to the boundary, the fact that $\widetilde{f} \circ \gamma = f_*(\gamma) \circ \widetilde{f}$ holds on $\mathbb{H}^n$ for all $\gamma \in \Gamma_1$ implies that it holds on the boundary, simply because $\mathbb{H}^n \subset \overline{\mathbb{H}^n}$ is dense.

The proof of the theorem involves some computations in hyperbolic geometry, and is broken up into the following lemmas.

Lemma 2

Let ℓ be a (geodesic) line in $\mathbb{H}^n$, and $\pi : \mathbb{H}^n \to \ell$ be the orthogonal projection onto ℓ. The maximum dilatation of π at $x \in \mathbb{H}^n$ is given by $d = \frac{1}{\cosh s}$, where $s = d(x, \ell)$.

Proof. This follows by some elementary geometry in the half-space model, see [Mar22] for the details. $\square$

Let $[x, y]$ denote the geodesic segment in $\mathbb{H}^n$ joining x and y.

Lemma 3

There is an $R > 0$ such that $F([p, q]) \subset N_R([F(p), F(q)])$, for any two distinct points p and q in $\mathbb{H}^n$.

Proof. Let C_1, C_2 be the pseudo-isometry constants of F, and fix big enough R such that $\cosh R > 2C_1^2$. For $p, q \in \mathbb{H}^n$, let ℓ be the line joining $F(p)$ and $F(q)$. We claim that $F([p,q])$ can only exit $N_R(\ell)$ for a small duration (thinking of time as parametrizing the geodesic $[p,q]$). Let $[r,s] \subset [p,q]$ be a maximal segment such that $F([r,s])$ is disjoint from the interior of ℓ. Being a pseudo-isometry gives,

$$\frac{1}{C_1} d(r,s) - C_2 \leq d(F(r), F(s)) \leq C_1 d(r,s).$$

Using the previous lemma, we can improve the upper bound to $C_1 \frac{d(r,s)}{\cosh R} + 2R$. Hence,

$$\left(\frac{1}{C_1} - \frac{C_1}{\cosh R} \right) d(r,s) \leq 2R + C_2.$$

As we took $\cosh R$ to be more than $2C_1^2$, we get $d(r,s) < M$ for some constant M depending only on C_1 & C_2, i.e, $F([p,q])$ exits $N_R(\ell)$ only on subsegments of length at most M. Using the fact that F is C_1-Lipschitz, we simply replace R with $R + C_1 M$ to get $F([p,q]) \subset N_{R+C_1M}(\ell)$. $\square$

The following lemma will give us the recipe to extend F to $\partial \mathbb{H}^n$.

Lemma 4

There is an $R > 0$ such that for all $p \in \mathbb{H}^n$ and half-line ℓ starting at p, there is a unique half-line ℓ' starting from $F(p)$ such that $F(\ell) \subset N_R(\ell')$.

Proof. We parametrize $\ell : [0, \infty) \to \mathbb{H}^n$ to have unit speed and $\ell(0) = p$. Since F is a pseudo-isometry, we get

$$\lim_{t \to \infty} d(F(p), F(\ell(t))) = \infty.$$

Let $v_t \in T_{F(p)}$ be the unit vector pointing to $F(\ell(t))$. It is easily checked that $\{v_t\}_{t \in \mathbb{N}}$ is a Cauchy sequence by looking at $R-$neighborhoods of $[F(p), F(\ell(t))]$, and hence must converge to some unit vector v based at $F(p)$. Let ℓ' be the half-line starting at $F(p)$ with direction given by this vector v. Then ℓ' satisfies the claim of the lemma. $\square$

Now, every point of $\partial \mathbb{H}^n$ is an equivalence class of half-lines starting at points in $\mathbb{H}^n$. We define $F : \partial \mathbb{H}^n \to \partial \mathbb{H}^n$ by sending the point representing some ℓ to the corresponding ℓ' constructed as in the preceding lemma. Unwrapping the definitions it is easy to check that this map is indeed well-defined and injective. To prove that the map is continuous, we need to improve the previous lemma to a result about lines, and then appeal to a geometric fact about pseudo-isometries.

Lemma 5

There is an $R > 0$ such that for each line $\ell \subset \mathbb{H}^n$, there is a unique line ℓ' such that $F(\ell) \subset N_R(\ell')$.

Proof. Fix some line ℓ and cut it into two half-lines. Using the last lemma, $F(\ell(t))$ converges to the distinct points $x_\pm \in \partial \mathbb{H}^n$ as $t \to \pm\infty$. Let ℓ' be the line joining these two points. ℓ' satisfies the claim, because for any $t > 0$ we have

$$F(\ell([-t,t])) \subset N_R([F(\ell(-t)), F(\ell(t))])$$

and we can let $t \to \infty$. $\square$

Lemma 6

There is an $R > 0$ such that for any line ℓ and hyperplane H orthogonal to ℓ, the image $F(H)$ projects orthogonally to ℓ' onto a bounded segment of length at most R. Here ℓ' is the line obtained as in the last lemma.

Proof. We first find R as in the previous lemma, and then find that F distorts hyperplanes orthogonal to ℓ only by some constant, depending only of the hyperplane chosen, by some geometric considerations. See [Mar22] for the details. $\square$

We finally finish this step of the proof now.

Lemma 7

The extended map $F : \overline{\mathbb{H}^n} \to \overline{\mathbb{H}^n}$ is continuous.

Proof. Consider some $x \in \partial\mathbb{H}^n$ and its image $F(x)$. Let ℓ be a half-line pointing to x, so that the half-line ℓ' obtained as in the preceding lemmas will point to $F(x)$. The set of half-spaces orthogonal to ℓ' form a neighborhood basis of $F(x)$; we fix some such half-space H.
Let $R > 0$ be as before, so that the image $F(\ell)$ is $R-$close to ℓ' and hence for sufficiently large t, $F(\ell(t))$ and its projection to ℓ' must lie in S at a distance at least R from the boundary ∂S. The previous lemma gives $F(P(t))$ is contained in H, if $P(t)$ is the hyperplane orthogonal to ℓ at $\ell(t)$. But this means that the half-space bounded by $P(t)$ is pushed inside H via F, for some t, i.e, F is continuous. $\square$

3 Second step of the proof

We now digress a little bit and look at ideal $n-$simplices in $\mathbb{H}^n$, i.e, geodesic $n-$simplices that have vertices at infinity. Let $\mathscr{S}_n$ denote the set of all such simplices.

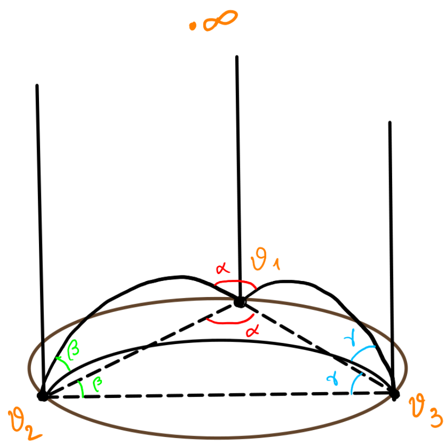
Definition 1 ▶ Regular simplices

A simplex in $\overline{\mathbb{H}^n}$ is said to be regular if any permutation of its vertices is induced by an isometry.

The definition above is the right analogue of the familiar notion of regular polyhedra in Euclidean geometry, and the following result ties the two notions together.

Lemma 8

Let $\sigma \in \mathscr{S}_n$ have vertices $\infty, v_1, \ldots, v_n$ where $v_i \in \mathbb{R}^n \times \{0\}$. Then σ is regular if and only if the Euclidean simplex on $v_1, \ldots, v_n$ is regular.



Proof. Let σ be regular. If λ is a permutation of $\{1, \ldots, n\}$, there exists $g \in \mathrm{Isom}(\mathbb{H}^n)$ such that $g(\infty) = \infty$ and $g(v_j) = v_{\lambda(j)}$ for all j. Using the explicit formulae of isometries of the upper half space model, it follows that $g\big|_{\mathbb{R}^{n-1}\times 0}$ must be an Euclidean isometry. Hence, the Euclidean simplex on $v_1, \ldots, v_n$ is regular.
Conversely, let the Euclidean simplex on $v_1, \ldots, v_n$ be regular. This of course means that every permutation of the vertices of σ that keeps ∞ fixed is induced by some isometry of $\mathbb{H}^n$. Moreover, for $1 \leq j \leq n$, each $v_i, i \neq j$ is at the same distance from v_j, say r. Hence, the inversion about the sphere of radius r centered at v_j exchanges the vertices v_j and ∞, keeping the rest fixed. This proves that σ is regular. $\square$

Theorem 6

The volume function vol() restricted to $\mathscr{S}_n$ attains its supremum v_n exactly at the regular and ideal $n-$simplices.

Before proving this theorem, we introduce some notation: an element of $\mathscr{S}_n$ will be denoted as $\tau[n]$, and the (unique up to isometry) regular ideal simplex will be denoted as $\tau_0[n]$. Similarly, let $\sigma[n]$ denote a geodesic Euclidean $n-$simplex, and $\sigma_0[n]$ a regular one.

For $n = 2$, every element of $\mathscr{S}_2$ is regular and has volume $v_2 = \pi$. For $n = 3$, it can be shown (the details are given in [Thu78] ch. 7, as well as [BP92] and [Mar22]) that

$$\mathrm{vol}(\tau[3]) = \Lambda(\alpha) + \Lambda(\beta) + \Lambda(\gamma),$$

where α, β, γ are the dihedral angles at any one vertex, also equal to the angles shown above, and Λ is the Lobachevsky function

$$\Lambda(\theta) = -\int_0^\theta \log|2\sin t|\mathrm{d}t.$$

Proof. (of Theorem (6) for $n = 3$)
Let $T(\tau[3])$ denote the similarity class of the Euclidean triangle associated to $\tau[3]$ as in Lemma (8). Then $\mathrm{vol}(\tau[3]) = 0$ iff $T(\tau[3])$ is degenerate. Hence, the theorem will be proven if show that the maximum of the map

$$(\alpha, \beta, \gamma) \mapsto \Lambda(\alpha) + \Lambda(\beta) + \Lambda(\gamma)$$

defined on the set $\{(\alpha, \beta, \gamma) \mid \alpha, \beta, \gamma > 0, \alpha + \beta + \gamma = \pi\}$ is attained exactly at the point $\left(\frac{\pi}{3}, \frac{\pi}{3}, \frac{\pi}{3}\right)$. This is clearly the same as showing

$$g(\alpha, \beta) = \Lambda(\alpha) + \Lambda(\beta) + \Lambda(\pi - \alpha - \beta)$$

attains its maximum exactly at $\left(\frac{\pi}{3}, \frac{\pi}{3}\right)$, when defined on $S = \{(\alpha, \beta) \mid \alpha, \beta > 0, \alpha + \beta < \pi\}$. But g is a positive continuous function that continuously extends to the boundary of S by 0, and hence has a maximum on S. It follows from the properties of Λ that g is of class C^1. The critical points of g are defined by

$$\begin{aligned} \Lambda'(\alpha) &= \Lambda'(\pi - \alpha - \beta) \\ \Lambda'(\beta) &= \Lambda'(\pi - \alpha - \beta), \end{aligned}$$

which is the same as

$$\begin{aligned} |\sin\alpha| &= |\sin(\pi - \alpha - \beta)| \\ |\sin\beta| &= |\sin(\pi - \alpha - \beta)|, \end{aligned}$$

and this system of equations has the unique solution $\left(\frac{\pi}{3}, \frac{\pi}{3}\right)$. □

For higher n, we need the following interlacing inequality.

Theorem 7

For $n \geq 2$,

$$\frac{n-1}{n^2} \leq \frac{\mathrm{vol}(\tau_0[n+1])}{\mathrm{vol}(\tau_0[n])} \leq \frac{1}{n}.$$

The upper bound was known to Thurston ([Thu78]), and the lower bound is due to Haagerup and Munkholm ([?]). The proof (whose details we omit here) involves manipulating the integrals defining the volume of $\tau_0[n]$ in the ball, half-space and projective model and is straightforward. Before completing the proof of Theorem (6), we need the following lemma.

Lemma 9

Let $f : (0,1] \to \mathbb{R}$ be a continuous concave function. If $\mathbf{z} \in \mathbb{R}^n$ is the center of mass of an arbitrary Euclidean simplex $\sigma[n]$ that has vertices on $\mathbb{S}^{n-1}$, and $z = \|z\|$, then

$$\mathrm{vol}(\sigma[n])^{-1} \int_{\sigma[n]} f(1-r^2)\mathrm{d}\mathbf{r} \leq \mathrm{vol}(\sigma_0[n])^{-1} \int_{\sigma_0[n]} f\big((1-z^2)(1-r^2)\big)\mathrm{d}\mathbf{r}$$

whenever the integrals converge. If f is strictly concave, equality holds iff σ is regular.

The proof again involves some straightforward analysis; see **[?]** for the details.

Proof. (of Theorem (6) for $n \geq 3$)
We induct on n, having already proven the result for $n = 3$. Consider an arbitrary $\tau[n+1]$ and set

$$f(t) = t^{-n/2} - K_n t^{-(n+1)/2}, \quad t \in (0,1],$$

where $K_n = \frac{n\,\mathrm{vol}(\tau_0[n+1])}{\mathrm{vol}(\tau[n])}$. It is easily checked that f is strictly concave on all of $(0,1]$ iff $K_n \geq \frac{n(n+1)}{(n+2)(n+3)}$. But we know $K_n \geq \frac{n-1}{n}$ because of Theorem (7), and this is more that $n(n+2)(n+3)$ if $n \geq 3$. The lemma above hence applies to f and the Euclidean simplex $\sigma[n]$ obtained from $\tau[n+1]$.
Let $\tau[n]$ be the simplex that maps to $\sigma[n]$ under the projective model. Then,

$$\begin{aligned}
n\,\mathrm{vol}(\tau[n+1]) - K_n\,\mathrm{vol}(\tau[n]) &\leq \int_{\sigma_0[n]} f((1-z^2)(1-r^2))\mathrm{d}\mathbf{r} \\
&= (1-z^2)^{-n/2} m\,\mathrm{vol}(\tau_0[n+1]) - K_n(1-z^2)^{-(n+1)/2}\,\mathrm{vol}(\tau_0[n]) \\
&\leq (1-z^2)^{-n/2}(n\,\mathrm{vol}(\tau_0[n+1]) - K_n\,\mathrm{vol}(\tau_0[n])) \\
&= 0.
\end{aligned}$$

The inductive hypothesis is $\mathrm{vol}(\tau[n]) \leq \mathrm{vol}(\tau_0[n])$ and so the above inequality now implies

$$n\,\mathrm{vol}(\tau[n+1]) \leq K_n\,\mathrm{vol}(\tau_0[n]) = n\,\mathrm{vol}(\tau_0[n+1]),$$

i.e, $\mathrm{vol}(\tau_0[n+1])$ is the supremum of vol() on $\mathscr{S}_{n+1}$. Moreover, the equality criterion of the lemma above proves that equality in the last inequality implies $\tau[n+1]$ is regular.

□

4 Gromov Norm

We now introduce the so-called *Gromov norm* or *simplicial volume* of manifolds, and prove Gromov's result that relates this quantity to the total volume in the case of hyperbolic manifolds. This is the key step in Gromov's proof of Mostow rigidity, and is also the most technical step involved.

First, consider an abstract topological space X and the $\mathbb{R}-$vector space of singular $k-$chains, $C_k(X;\mathbb{R})$. We make this into a normed vector space by defining

$$\|c\| = \inf\left\{\sum_i |a_i| \;\middle|\; c = \sum_i a_i\sigma_i\right\},$$

for each $c \in C_k(X;\mathbb{R})$. This descends to a semi-norm on the homology group $H_k(X;\mathbb{R}) = Z_k(X)/B_k(X)$ by defining

$$\|z\| = \inf\{\|c\| \mid c \in Z_k, z = [c]\},$$

for each $z \in H_k(X;\mathbb{R})$. This is not a norm in general, because non-zero elements in $B_k(X)$ are assigned zero value by this map.

Now consider a compact manifold M of dimension n, not assumed to be hyperbolic. It is a general fact from homology that $H_n(M;\mathbb{Z}) \simeq \mathbb{Z}$, and the same is true if we take real coefficients. Moreover, we have a distinguished generator $[M] \in H_n(M)$, called the *fundamental class* of M.

Definition 2 ▶ Gromov Norm

For a compact manifold M of dimension n, with fundamental class $[M] \in H_n(M;\mathbb{R})$, we define its Gromov norm to be $\|M\| = \|[M]\|$.

Before discussing Gromov's theorem, we deduce some some easy properties of the Gromov norm.

(i) If $f : M \to N$ is a continuous map of compact manifolds, then $\|M\| \geq |\deg f|\|N\|$. In particular, $\|\cdot\|$ is a homotopy invariant.

Proof. If $z \in H_k(M)$, and f_* denotes the map induced by f on homology, we get

$$\|f_*(z)\| \leq \|z\|,$$

because if $z = [\sum a_i\sigma_i]$, $f_*(z) = [\sum a_i(f \circ \sigma_i)]$. Further, the degree of f satisfies $f_*([M]) = \deg f \cdot [N]$. Taking f to be a homotopy equivalence, we get $\|\cdot\|$ is homotopy invariant because homology is. □

(ii) If M admits a continuous self-map of degree at least 2 (in absolute value), then $\|M\| = 0$. Hence, all spheres and the torus have zero Gromov norm.

Proof. This assertion is immediate from the last one. The second part follows from the fact that such maps exist if the manifold is a sphere or the torus. □

In light of the second property above, it is not a priori obvious that there even exist manifolds that have nonzero Gromov norm. However, Gromov's theorem provides us with an abundant class of examples.

Theorem 8 (Gromov)

Let M be a compact oriented hyperbolic manifold. Then $\mathrm{vol}(M) = v_n\|M\|$.

Corollary : All compact hyperbolic manifolds have non-zero Gromov norm, and hyperbolic volume is a homotopy invariant. Moreover, if f is a continuous self-map of such a manifold, $|\deg f| \leq 1$.

Using the corollary above, we can think of Gromov's theorem as a weaker rigidity result: hyperbolic volume is a topological invariant. Gromov's proof of Mostow rigidity really shows that this implies that the metric itself must also be a topological invariant.

The proof of Gromov's theorem is broken up into the two inequalities $\mathrm{vol}(M) \leq v_n\|M\|$ and $\mathrm{vol}(M) \geq v_n\|M\|$. The first inequality is relatively easy to prove, but the second one requires some work. The rest of this section is devoted to the proof of these two inequalities; we fix a compact oriented hyperbolic $n-$manifold M with universal cover $p : \mathbb{H}^n \to M$.

4.1 Proof of $\mathrm{vol}(M) \leq v_n\|M\|$

For distinct points $u_0, \ldots, u_n \in \mathbb{H}^n$, we denote the (barycentric parametrization of the) geodesic $n-$simplex in $\mathbb{H}^n$ on these vertices by $\sigma(u_0, \ldots, u_n)$.

Definition 3 ▶ Straight $n-$chains

We call a singular $n-$simplex $\varphi : \Delta_n \to M$ *straight* if $\varphi = p \circ \sigma(u_0, \ldots, u_n)$ for some u_j's. A singular $n-$chain is *straight* if it can be expressed as a linear combination of straight $n-$simplices.

We also say that $\sigma(u_0, \ldots, u_n)$ is *degenerate* if the u_j's are contained in some hyperbolic $(n-1)-$subspace; this is equivalent to saying $\mathrm{vol}(\sigma(\Delta_n)) = 0$. Any such simplex is homologous to a sum of non-degenerate ones, and so we do not consider such simplices in the sequel. From now, we mean $n-$chains and $n-$simplices if not specified otherwise.

Lemma 10 (Chain straightening)

Every singular chain in M is naturally homotopic (and hence homologous) to a straight one.

Proof. Let $\varphi : \Delta_n \to M$ be an arbitrary $n-$simplex and consider the lift $\widetilde{\varphi} : \Delta_n \to \mathbb{H}^n$, which exists because the standard simplex is simply connected. Let $v_j = \widetilde{\varphi}(e_j), 0 \leq j \leq n$, and consider $\varphi^{\text{st}} = p \circ \sigma(v_0, \dots, v_n)$. We have a homotopy F from φ^{st} to φ given by

$$F(z,t) = p(t \cdot \widetilde{\varphi}(z) + (1-t) \cdot \sigma(v_0, \dots, v_n)(z)), \ (z,t) \in \Delta_n \times [0,1].$$

F satisfies $F(e_j, s) = \varphi(e_j)$ for all $s \in [0,1]$. Hence, if $\sum a_i \varphi_i$ is singular chain, $\sum a_i \varphi_i^{\text{st}}$ represents the same homology class. $\square$

Using the above lemma, we are reduced to showing the following: if $\sum a_i \sigma_i$ is a straight cycle representing $[M]$, then

$$\sum |a_i| \geq \frac{\text{vol}(M)}{v_n}.$$

To accomplish this, we define the following signed analog of volume.

Definition 4 ▶ Algebraic Volume

If $\varphi = p \circ \sigma$ is straight simplex, we define the *algebraic volume* of φ to be

$$\text{algvol}(\varphi) = \begin{cases} \text{vol}(\sigma(\Delta_n)), & \text{d}_t\varphi \text{ is orientation preserving} \\ -\text{vol}(\sigma(\Delta_n)), & \text{otherwise.} \end{cases}$$

We note that since p is a local isometry and σ is non-degenerate, $\text{d}_t\varphi$ either preserves or reverses orientation at every $t \in \text{int}(\Delta_n)$. For future reference, we can also express the algebraic volume as

$$\text{algvol}(\varphi) = \int_{\varphi(\Delta_n)} \alpha(x) \, \text{d}\,\text{vol}(x) \,,$$

where $\alpha(x) = \alpha_+(x) - \alpha_-(x)$, and

$$\alpha^+(x) = |\{t \in \text{int}(\Delta_n) \mid \varphi(t) = x, \text{d}_t\varphi \text{ is orientation preserving}\}|$$
$$\alpha^-(x) = |\{t \in \text{int}(\Delta_n) \mid \varphi(t) = x, \text{d}_t\varphi \text{ is orientation reversing}\}|.$$

Of course, by the preceding remark we know $\alpha^+(x) \neq 0$ for some x implies $\alpha^-(x) = 0$ for all x, and hence $\alpha^+ \cdot \alpha^- = 0$ always. We also extend the definition of algebraic volume to straight chains by linearity, adopting the convention that all chains are written in the shortest length expression. Finally, we also note

$$|\text{algvol}(\varphi)| = \text{vol}(\sigma(\Delta_n)) \leq v_n,$$

for any straight simplex φ, by the definition of v_n.

Theorem 9

Let $z = \sum_{i=1}^k a_i \varphi_i$ be a straight cycle representing $[M]$. Then,

$$\text{vol}(M) \leq v_n \sum_{i=1}^k |a_i|.$$

This theorem proves that $\text{vol}(M) \leq v_n \|M\|$ because of Lemma (10) as mentioned before.

Proof. Let $N = \cup \varphi_i(\partial \Delta_n)$. We define $\alpha_i^{\pm}(x)$ for each i as before, and set

$$\Phi_z(x) = \sum_{i=1}^k a_i \alpha_i(x).$$

We claim that $\Phi_z(x) = 1$ if $x \in M \setminus N$. Fix some triangulation of M, and straighten the simplices to get a representative z_0 of $[M]$, which is also a straight simplex. We define Φ_{z_0} in the same fashion, and note that $\Phi_{z_0}(x) = 1$ if x does not lie in the $(n-1)-$skeleton of the triangulation. We now fix $x \in M \setminus N$. By definition, Φ_z is locally constant on $M \setminus N$ and N is a closed subset, and so we can assume without loss of generality that x is not in the $(n-1)-$skeleton. This means we are reduced to proving

$$\Phi_{z_0-z}(x) = \Phi_{z_0}(x) - \Phi_z(x) = 0.$$

Recall that the inclusion $i : (M, \emptyset) \hookrightarrow (M, M \setminus \{x\})$ induces an isomorphism

$$i_* : H_n(M) \to H_n(M, M \setminus \{x\}).$$

We get $i_*([w] = \Phi_w(x))$ if $w = \sum b_j w_j$ is straight and $x \notin \cup w_j(\partial\Delta_n)$. This then implies $\Phi_{z-z_0}(x) = 0$, as was claimed.

Now, N has zero volume and so

$$\mathrm{vol}(M) = \int_M \Phi(x)\, \mathrm{d}\,\mathrm{vol}(x)\,,$$

because we just showed $\Phi \equiv 1$ outside N. We alo have,

$$\mathrm{algvol}(\varphi_i) = \int_{\varphi_i(\Delta_n)} \alpha_i(x)\, \mathrm{d}\,\mathrm{vol}(x) = \int_M \alpha_i(x)\, \mathrm{d}\,\mathrm{vol}(x)\,.$$

Combining these two equalities we get, $\mathrm{vol}(M) = \sum a_i\, \mathrm{algvol}(\varphi_i)$. Hence, by the final remark before this theorem on v_n,

$$\mathrm{vol}(M) = \left|\sum a_i\, \mathrm{algvol}(\varphi_i)\right| \le \sum |a_i||\mathrm{algvol}(\varphi_i)| \le v_n \sum |a_i|,$$

as required. $\square$

4.2 Proof of $\mathrm{vol}(M) \ge v_n \|M\|$

We need another definition before starting with the proof of this direction.

Definition 5 ▶ $\varepsilon-$efficient cycles

A straight cycle $\varphi = \sum a_i \varphi_i$ that represents $[M]$ is called $\varepsilon-$*efficient* if

$$\mathrm{sgn}(a_i) \cdot \mathrm{algvol}\, \varphi_i \ge v_n - \varepsilon,$$

holds for all i, where $\varepsilon > 0$.

It is clear that if we can produce an $\varepsilon-$efficient cycle for each $\varepsilon > 0$, we could conclude $\mathrm{vol}(M) \ge v_n\|M\|$ as we have already shown in the last section that $\mathrm{vol}(M)$ is equal to the algebraic volume of any straight cycle representing $\|M\|$. We now proceed to construct such efficient cycles.

Recall that $\mathrm{Isom}(\mathbb{H}^n) = SO(n,1)$ is a compact Lie group and hence admits a bi-invariant Haar measure μ, which is unique up to scaling. We also need the following geometric fact about simplices in $\mathbb{H}^n$.

Lemma 11

Let $u_0, \ldots, u_n \in \mathbb{H}^n$ and σ be their convex hull. Then σ is regular iff $d(u_i, u_j)$ is the same constant for all $i \ne j$.

We omit the proof of this result; it follows from some elementary geometric considerations as in Lemma (8). See [BP92] for the details. For $R > 0$, we set

$$\mathscr{S}(R) = \{(u_0, \ldots, u_n) \in (\mathbb{H}^n)^{n+1} \mid d(u_i, u_j) = R\, \forall\, i \ne j\}.$$

We also fix some point $(u_0^R, \ldots, u_n^R) \in \mathscr{S}(R)$ for each $R > 0$.

Lemma 12

The map

$$\begin{aligned}\psi : \mathrm{Isom}(\mathbb{H}^n) &\to \mathscr{S}(R)\\ g &\mapsto \big(g(u_0^R), \ldots, g(u_n^R)\big)\end{aligned}$$

is a bijection.

Proof. Let $\psi(g_1) = \psi(g_2)$. We use the ball model and assume $u_0^R = 0$. This means that $g_1^{-1} \circ g_2$ is a linear map fixing n linearly independent points fixed and hence must be the identity. This proves injectivity.

Let $(u_0, \ldots, u_n) \in \mathscr{S}(R)$. Applying a suitable isometry, we can assume $u_0 = u_0^R = 0$ (again in the ball model). Since both u_1 and u_1^R lie on the sphere of radius R centered at 0, there is $A \in O(n)$ such that $Au_1^R = u_1$, and so we reduce to $u_1^R = u_1$ by applying the corresponding isometry. We proceed in this fashion and conclude that ψ is surjective. $\square$

Using this bijection the Haar measure induces a measure on $\mathscr{S}(R)$ defined by

$$m(A) = \mu(\{g \in \mathrm{Isom}(\mathbb{H}^n) \mid \psi(g) \in A\}),$$

for any $A \subset \mathscr{S}(R)$; here we of course mean A is measurable iff $\psi^{-1}(A)$ is. The right invariance of μ implies that the choice of $(u_0^R, \ldots, u_n^R)$ is irrelevant.

An element $(u_0, \ldots, u_n) \in \mathscr{S}(R)$ is called *positive* if $\sigma(u_0, \ldots, u_n)$ is orientation preserving at all $z \in \Delta_n$. This splits up $\mathscr{S}(R)$ into $\mathscr{S}_\pm(R)$, and if $(u_0^R, \ldots, u_n^R)$ is positive then $\mathscr{S}_+(R)$ corresponds to the orientation preserving isometries of $\mathbb{H}^n$ under ψ.

We also set $\widetilde{\mathscr{S}}(R) = \{\sigma(u_0, \ldots, u_n) \mid (u_0, \ldots, u_n) \in \mathscr{S}(R)\}$. Note that if $\tau, \tau' \in \widetilde{\mathscr{S}}(R)$, there $g \in \mathrm{Isom}(\mathbb{H}^n)$ such that $g(\tau) = \tau'$ and so $\mathrm{vol}(\tau) = \mathrm{vol}(\tau')$. This means $V(R) = \mathrm{vol}(\tau), \tau \in \widetilde{\mathscr{S}}(R)$ is a well-defined map V. We need one last result before beginning the actual construction of the efficient cycles.

Lemma 13

$\lim_{R\to\infty} V(R)$ exists and is equal to v_n.

Proof. Clearly V is a non-decreasing function that is bounded above by v_n. But we can find a mapping $R \mapsto \tau_R$ such that τ_R “increase” to a regular ideal simplex τ so that $\lim \mathrm{vol}(\tau_R) = \mathrm{vol}(\tau) = v_n$. $\square$

Let $\Gamma \simeq \pi_1(M)$ so that the covering $p : \mathbb{H}^n \to M$ is given by quotient by the action of Γ. Consider the action of Γ on Γ^{n+1} by left translations

$$(\gamma_0, \ldots, \gamma_{n+1}) \xrightarrow{\gamma} (\gamma \cdot \gamma_0, \ldots, \gamma \cdot \gamma_{n+1}),$$

and let $\Omega = \Gamma^{n+1}/\Gamma$ be the quotient space of this action.

We also fix a compact Dirichlet domain $D \subset \mathbb{H}^n$ for Γ, with diameter d, and also fix a point $u \in \mathrm{int}(D)$. For $\omega = [(\gamma_0, \ldots, \gamma_n)] \in \Omega$, we define the simplex

$$\sigma_\omega = p \circ \sigma(\gamma_0(u), \ldots, \gamma_n(u)).$$

We also define (for $R > 0$),

$$a_R^+(\omega) = m(\{(u_0, \ldots, u_n) \in \mathscr{S}_+(R) \mid u_j \in \gamma_j(D) \,\forall\, j\}).$$

It is easy to check that this is well-defined unwrapping the definitions of the measure m and the simplex ω. Since we chose D to be compact, $\{\delta \in \mathrm{Isom}(\mathbb{H}^n) \mid \delta(x_0) \in D\}$ is of finite measure for any x_0, and hence $a_R^+(\omega) < \infty$. Replacing $\mathscr{S}_+(R)$ with $\mathscr{S}_-(R)$, we define $a_R^-(\omega)$ analogously. Finally, we consider

$$a_R(\omega) = a_R^+(\omega) - a_R^-(\omega),$$

and define the formal sum

$$z_R = \sum_{\omega \in \Omega} a_R(\omega)\sigma_\omega.$$

We will now proceed with the proof (in a series of steps) that z_R is a cycle homologous to $[M]$, and will be efficient for suitable R.

(i) z_R is expressed by a finite sum.

Note that each ω has a unique representative $(1, \gamma_1, \dots, \gamma_n)$. If $a_R(\omega) \neq 0$, there is at least one simplex $\sigma(u_0, \dots, u_n) \in \widetilde{\mathscr{S}}(R)$ such that

$$u_0 \in D \quad u_i \in \gamma_i(D), 1 \leq i \leq n.$$

We then get $d(u, \gamma_i(u)) \leq 2d + R$ for $1 \leq i \leq n$. But now, Γ being discrete implies that there are only finitely many choices for the γ_i's. Hence, $a_R(\omega) \neq 0$ only for finitely many ω's.

(ii) z_R is a cycle.

We claim that for any $(n-1)-$face τ of a σ_ω that appears in z_R, the corresponding coefficient in ∂z_R is zero. Fix some such τ, which by construction is obtained as the projection of a geodesic $(n-1)-$simplex having vertices in the $\Gamma-$orbit of u, i.e,

$$\tau(t_0, \dots, t_{n-1}) = p\Big(\sum t_j \gamma_j(u)\Big),\ (t_0, \dots, t_{n-1}) \in \Delta_{n-1},$$

for suitable $\gamma_0, \dots, \gamma_{n-1}$. The coefficient corresponding to τ in ∂z_R is then exactly

$$\sum_{j=0}^{n} (-1)^{n-j} \sum_{\gamma \in \Gamma} a_R([(\gamma_0, \dots, \gamma_{j-1}, \gamma, \gamma_{j+1}, \dots, \gamma_{n-1})]).$$

We show that the inner sum is 0 for each j; without loss of generality we take $j = n$, for which this inner sum is

$$\sum_{\gamma_n \in \Gamma} a_R([(\gamma_0, \dots, \gamma_n)]) = \sum_{\gamma_n \in \Gamma} a_R^+([(\gamma_0, \dots, \gamma_n)]) - \sum_{\gamma_n \in \Gamma} a_R^-([(\gamma_0, \dots, \gamma_n)]).$$

Now, using the definition of a_R^+ and a_R^- in terms of the measure m, we get the summands on the right side are exactly $m(A_+)$ and $m(A_-)$ respectively, where

$$A_\pm = \{(u_0, \dots, u_n) \in \mathscr{S}_\pm(R) \mid u_i \in \gamma_i(D)\, \forall\, i \leq n-1\}.$$

But if $g_0 \in \operatorname{Isom}(\mathbb{H}^n)$ is the reflection about the hyperbolic hyperplane containing the u_i's we have

$$\big\{g \in \operatorname{Isom}^-(\mathbb{H}^n) \mid g(u_i^R) \in \gamma_i(D)\, \forall\, i < n\big\} = \big\{g \in \operatorname{Isom}^+(\mathbb{H}^n) \mid g(u_i^R) \in \gamma_i(D)\, \forall\, i < n\big\} \cdot g_0,$$

and so $m(A_-) = m(A_+)$. This completes the proof that z_R is a cycle.

(iii) If $R > 2d$, then $a_R^+(\omega) \cdot a_R^-(\omega) = 0$ for all ω.

If $\sigma_0 \in \widetilde{\mathscr{S}}(R)$ has the first vertex in D, the second in $\gamma_1(D)$, …and the n^{th} vertex in $\gamma_n(D)$, then any other element of $\widetilde{\mathscr{S}}(R)$ with this property must have the same orientation as σ_0. So we get $a_R^+(\omega) = 0 \iff a_R^-(\omega) \neq 0$.

(iv) For all $\varepsilon > 0$, if R is big enough then $|\operatorname{algvol}(\sigma_\omega)| \geq v_n - \varepsilon$ whenever $a_R(\omega) \neq 0$.

Suppose $\omega = [(\gamma_0, \dots, \gamma_n)]$ satisfies $a_R(\omega) \neq 0$. This means that there is $\sigma(u_0, \dots, u_n) \in \widetilde{\mathscr{S}}(R)$ such that $u_i \in \gamma_i(D)$ and thus $d(\gamma_i(u), u_i) \leq d$. Further, $|\operatorname{algvol}(\sigma_\omega)| = \operatorname{vol}(\sigma(\gamma_0(u), \dots, \gamma_n(u)))$. But we now note that $\sigma(\gamma_0(u), \dots, \gamma_n(u))$ has vertices that are distance at most d away from an element of $\widetilde{\mathscr{S}}(R)$, Lemma (13) implies that this volume converges to v_n as $R \to \infty$.

(v) If $R > 2d$ and $a_R(\omega) \neq 0$, then $a_R(\omega) \cdot \mathrm{algvol}(\sigma_\omega) > 0$.

Suppose $a_R(\omega) \neq 0$. We define, for $x \in M$, $\alpha_\omega(x)$ as before:

$$|\{t \in \mathrm{int}(\Delta_n) \mid \sigma_\omega(t) = x, \mathrm{d}_t\sigma_\omega \text{ positive}\}| - |\{t \in \mathrm{int}(\Delta_n) \mid \sigma_\omega(t) = x, \mathrm{d}_t\sigma_\omega \text{ negative}\}|.$$

We again have

$$\mathrm{algvol}(\sigma_\omega) = \int_M \alpha_\omega(x)\, \mathrm{d}\,\mathrm{vol}(x)\,,$$

and so we only need to show $a_R(\omega)\alpha_\omega(x) \geq 0$.

We claim that $\alpha_\omega(x) > 0$ if it is non-zero and $a_R^+(\omega) \neq 0$, and the same would hold if $a_R^-(\omega) \neq 0$ of course. Let $p(\widetilde{x}) = x$, and consider the lift $\widetilde{\sigma}_\omega$ of σ_ω starting at $\widetilde{x}$, which is given by $\sigma(\gamma_0(u), \ldots, \gamma_n(u))$ with $\omega = [(\gamma_0, \ldots, \gamma_n)]$. If $a_R^+(\omega) \neq 0$, we can find $(u_0, \ldots, u_n) \in \mathscr{S}_+(R)$ such that $u_i \in \gamma_i(D)$, and thus $d(u_i, \gamma_i(u)) \leq d$. This then implies $\sigma(\gamma_0(u), \ldots, \gamma_n(u))$ is positively oriented and hence $\alpha_\omega(x) > 0$.

(vi) If $R > 2d$ then $\mathrm{algvol}(z_R) > 0$, so that in particular z_R is a non-trivial cycle.

We claim that $a_R(\omega)$ is non-vanishing for some $\omega \in \Omega$, so that (v) implies the result. Pick $(u_0^0, \ldots, u_n^0) \in \mathscr{S}(R)$; since $\Gamma(D) = \mathbb{H}^n$, we can find $\gamma_0, \ldots, \gamma_n \in \Gamma$ such that $u_i^0 \in \gamma_i(D)$ for all i. By perturbing the u_i^0's slighly and choosing γ_i's for these, we can assume $u_i^0 \in \gamma_i(\mathrm{int}(D))$ for all i. We thus have

$$m(\{(u_0, \ldots, u_n) \in \mathscr{S}(R) \mid u_i \in \gamma_i(D)\,\forall\, i\}) \neq 0,$$

and hence, for $\omega = [(\gamma_0, \ldots, \gamma_m)]$ either $a_R^+(\omega) \neq 0$ or $a_R^-(\omega) \neq 0$, which then implies $a_R(\omega) \neq 0$.

Proof. (Conclusion of the proof that $\mathrm{vol}(M) \geq v_n\|M\|$)
Given $\varepsilon > 0$, we take $R > 2d$ so that (iv) applies. Then, by (i) and (ii), z_R is a singular cycle. By (iv) and (v),

$$\mathrm{sgn}(a_R(\omega)) \cdot \mathrm{algvol}(\sigma_\omega) \geq v_n - \varepsilon$$

holds whenever $a_R(\omega) \neq 0$.
By (vi), there is $k \neq 0$ we have $[z_R] = k[M]$, i.e, $\frac{1}{k}z_R$ represents $[M]$, so that $\mathrm{algvol}(z_R) = k\,\mathrm{vol}(M)$. In particular, $k > 0$ and so

$$\frac{1}{k}z_R = \sum_{\omega \in \Omega} \frac{a_R(\omega)}{k}\sigma_\omega$$

is a straight cycle such that

$$\mathrm{sgn}\left(\frac{a_R(\omega)}{k}\right) \cdot \mathrm{algvol}(\sigma_\omega) \geq v_n - \varepsilon.$$

This exactly means that $\frac{1}{k}z_R$ is an $\varepsilon-$efficient cycle, which completes the proof. ☐

5 Third step of the proof

Theorem 10

Let $\widetilde{f} : \overline{\mathbb{H}}^n \to \overline{\mathbb{H}}^n$ be as in the first step. Then, if $\{u_0, \ldots, u_n\}$ are the vertices of a simplex of volume v_n, the simplex on $\left\{\widetilde{f}(u_0), \ldots, \widetilde{f}(u_n)\right\}$ also has volume v_n.

Proof. Assume towards a contradiction that there exists some simplex $\tau = \sigma(w_0, \ldots, w_n)$ in $\overline{\mathbb{H}^n}$ such that $\mathrm{vol}(\tau) = v_n$ and

$$\mathrm{vol}(\sigma(\widetilde{f}(w_0), \ldots, \widetilde{f}(w_n))) = v_n - 2\varepsilon < v_n.$$

Without loss of generality, τ is positively oriented. As $\widetilde{f}$ is continuous, there are neighborhoods $U_j \subset \overline{\mathbb{H}^n}$ of w_j such that if $u_j \in U_j$, then

$$\mathrm{vol}(\sigma(\widetilde{f}(w_0), \dots, \widetilde{f}(w_n))) = v_n - \varepsilon \leq v_n.$$

We fix the notation of the proof that $\mathrm{vol}(M) \geq v_n \|M\|$ for $M = M_1, p = p_1, \Gamma = \Gamma_1$. We define

$$c_R = \sum \{a_R(\omega)\sigma_\omega \mid \omega = [(\gamma_0, \dots, \gamma_n)] \in \Omega, \gamma_i(u) \in U_i \, \forall \, i\}.$$

We assume the following lemma (the proof involves computations similar to the proof of Gromov's theorem; see [BP92] for the details) and finish the proof.

Lemma 14

There exist $\alpha_1, \alpha_2 > 0$ such that if R is sufficiently big then $\|z_R\| = \alpha_1$ and $\|c_R\| \geq \alpha_2$.

Since f is a homotopy equivalence, we have $f_*([M_1]) = \pm[M_2]$, and so $\|M_1\| = \|M_2\|$ and $\mathrm{vol}(M_1) = \mathrm{vol}(M_2)$, because of Gromov's theorem. We also recall there is $k > 0$ such that $z_R = k[M_1]$, so that $f_*([z_R]) = \pm k[M_2]$. A representative of $f_*([z_R])$ is given by the straightening z'_R of the cycle $f \circ z_R$:

$$z'_R = \sum_{\omega \in \Omega} a_R(\omega) \Big(p_2 \circ \sigma(\widetilde{f}(\gamma_0(u)), \dots, \widetilde{f}(\gamma_n(u))) \Big).$$

Now, $[z_R] = k[M_1]$ and $[z'_R] = \pm k[M_2]$ gives $\mathrm{algvol}(z_R) = k \, \mathrm{vol}(M_1)$ and $\mathrm{algvol}(z'_R) = \pm k \, \mathrm{vol}(M_2)$. But the volumes are equal by Gromov's theorem and so,

$$\mathrm{algvol}(z_R) = \pm \, \mathrm{algvol}(z'_R).$$

We also recall that if $a_R(\omega) \neq 0$, for some $\omega = [(\gamma_0, \dots, \gamma_n)]$, then the volume of $\sigma(\gamma_0(u), \dots, \gamma_n(u))$ differs from v_n by a quantity that goes to 0 as $R \to \infty$. Hence,

$$|\mathrm{algvol}(z_R)| \geq \|z_R\| \inf \{\mathrm{vol}(\sigma(\gamma_0(u), \dots, \gamma_n(u))) \mid a_R(\omega) \neq 0\},$$

and the right side converges to $\alpha_1 v_n$.
We also have,

$$\begin{aligned}
\big|\mathrm{algvol}(z'_R)\big| &= \sum_{\omega \in \Omega} |a_R(\omega)| \cdot \mathrm{vol}\big(\sigma(\widetilde{f}(\gamma_0(u)), \dots, \widetilde{f}(\gamma_n(u)))\big) \\
&= \sum \Big\{ |a_R(\omega)| \cdot \mathrm{vol}\big(\sigma(\widetilde{f}(\gamma_0(u)), \dots, \widetilde{f}(\gamma_n(u)))\big) \mid \gamma_i(u) \notin U_i \text{ for some } i \Big\} \\
&+ \sum \Big\{ |a_R(\omega)| \cdot \mathrm{vol}\big(\sigma(\widetilde{f}(\gamma_0(u)), \dots, \widetilde{f}(\gamma_n(u)))\big) \mid \gamma_i(u) \in U_i \text{ for all } i \Big\} \\
&\leq v_n \sum \{|a_R(\omega)| \mid \gamma_i(u) \notin U_i \text{ for some } i\} + (v_n - \varepsilon) \sum \{|a_R(\omega)| \mid \gamma_i(u) \in U_i \text{ for all } i\} \\
&\leq v_n(\|z_R\| - \|c_R\|) + (v_n - \varepsilon)\|c_R\| \\
&\leq \alpha_1 v_n \left(1 - \frac{\varepsilon \alpha_1}{v_n \alpha_2}\right)
\end{aligned}$$

But this is a contradiction as

$$\liminf \big|\mathrm{algvol}(z'_R)\big| = \liminf |\mathrm{algvol}(z_R)| \geq \alpha_1 v_n.$$

This concludes the third step of the proof. □

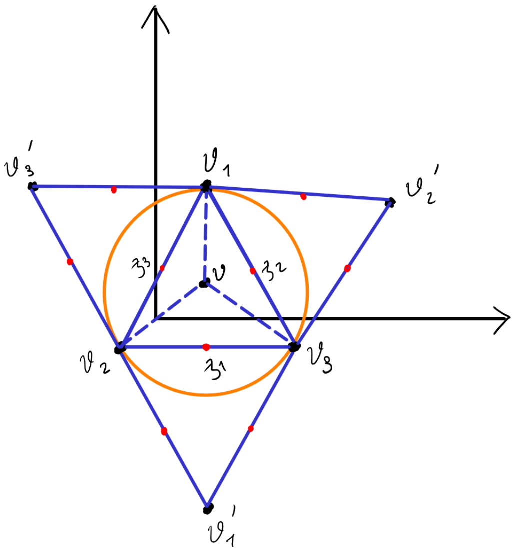
6 Fourth step of the proof

Finally, we prove the following general fact about pseudo-isometries that ties all the steps above to complete the proof that $\widetilde{f}$ will be induced by an isometry.

Theorem 11

Let $n \geq 3$ and $P : \partial\mathbb{H}^n \to \partial\mathbb{H}^n$ be a continuous injection, such that $\mathrm{vol}(\sigma(P(u_0), \ldots, P(u_n))) = v_n$ whenever $\mathrm{vol}(\sigma(u_0, \ldots, u_n)) = v_n$. Then P is induced by an isometry.

Proof. In light of the second step, the condition on P is equivalent to saying that it sends vertices of any maximal volume simplex to the vertices of another one. But we also know that the maximal volume simplices are the regular ideal ones, and are in particular all isometric. Hence, there is an isometry $Q \in \mathrm{Isom}(\mathbb{H}^n)$ so that $P' := P \circ Q$ fixes the regular ideal simplex with vertices $\infty, v_1, \ldots, v_n$, where v_j lie on the boundary plane $\mathbb{R}^n \times 0$. Using lemma (8) again, we get the Euclidean simplex on $v_1, \ldots, v_n$ is regular. By some elementary geometry, we know that a regular $n-$simplex tiles $\mathbb{R}^n$. The following figure depicts the scene for $n = 3$:



We now proceed to analyse the action of $P \circ Q$ on the plane $\mathbb{R}^n \times 0$; for the sake of clarity and brevity of notation we do the analysis in dimension 3, but the arguments go through verbatim for higher n. Consider the reflection v_1' of v_1 about the line through v_2 and v_3. The simplex on v_1', v_2, v_3, ∞ is regular and P' fixes v_2, v_3, ∞, and hence we get $P'(v_1') = v_1'$ simply because there are no other choices for the image; given any two points on the plane there are exactly two choices for the third vertex so that the simplex formed is regular. The same applies to v_2' and v_3' and repeating the arguments, we find that P' must fix all the vertices of the tesselation of the plane obtained from the regular simplex on v_1, v_2, v_3.

Now consider the center v of the simplex on v_1, v_2, v_3. The inversion centered at v induces an isometry of $\mathbb{H}^n$ that fixes the v_j's and maps ∞ to v, and so we get $P'(v)$ is either v or ∞. As P' fixes ∞, we must have $P'(v) = v$. But now, the midpoint z_1 of the segment joining v_2 and v_3 is the inversion of v_1' about v, and hence must again be fixed by the same reasoning. Therefore, P' fixes the midpoints of each segment of the simplices that are present in the tesselation we started with, and repeating this process now for the new tesselation obtained from the simplex on the midpoints, we get P' must fix a dense set of points. By continuity, P' must be the identity map, i.e, $P = Q^{-1}$ is an isometry. $\square$

7 End of Gromov's proof

Let us take stock of what we've shown so far: starting with a homotopy equivalence $f : M_1 \to M_2$, we've shown that the lift $\widetilde{f} : \mathbb{H}^n \to \mathbb{H}^n$ extends to all of $\overline{\mathbb{H}^n}$ continuously, and injects $\partial\mathbb{H}^n$ to itself. Moreover, we have shown that $\widetilde{f}$ maps the vertices of one regular ideal simplex to the vertices of another one. But now the result of the last step can be applied to conclude that there is an isometry $Q \in \mathrm{Isom}(\mathbb{H}^n)$ such that $\widetilde{f}$ agrees with Q on $\partial\mathbb{H}^n$. Also recall that the relation

$$\widetilde{f} \circ \gamma = f_*(\gamma) \circ \widetilde{f}$$

holds for each element $\gamma \in \Gamma_1$ on all of $\overline{\mathbb{H}^n}$. In particular this means that on $\partial\mathbb{H}^n$,

$$Q \circ \gamma = f_*(\gamma) \circ Q$$

holds for all $\gamma \in \Gamma_1$. But now this is an equality involving only isometries of $\mathbb{H}^n$, and it is a general fact that if such an equality holds on the boundary at infinity, it must in fact hold on all of $\mathbb{H}^n$. We now define the candidate isometry $q : M_1 \to M_2$ as

$$q(p_1(x)) = p_2(Q(x)), x \in \mathbb{H}^n.$$

Here p_i denotes the covering map $\mathbb{H}^n \to M_i$, given by quotienting by the action of Γ_i. Because of the relation above of Q with the action of Γ_1, this is a well-defined bijection, and it is an isometry because p_1, p_2 are local isometries as well. Finally,

$$H(t, p_1(x)) = p_2(t\widetilde{f}(x) + (1-t)Q(x)), x \in \mathbb{H}^n, t \in [0,1],$$

defines a homotopy from q to f. This concludes the proof of Mostow rigidity (Theorem (1)). ■

8 An Application of Mostow Rigidity

Let S_g be the closed oriented surface of genus g, and endow it with any hyperbolic metric. We recall the following two classical results.

Theorem 12 (Dehn-Nielsen-Baer)

$\mathrm{Out}(\pi_1(S_g))$ is isomorphic to $\mathrm{Mod}(S_g)$, and in particular is an infinite group.

Theorem 13 (Hurwitz)

$\mathrm{Isom}(S_g)$ has size at most $84(g-1)$.

Using Mostow rigidity, we prove the following result which shows the behaviour above is specific to dimension 2.

Theorem 14

Let M be a closed oriented hyperbolic manifold of dimension $n \geq 3$. Then $\mathrm{Out}(\pi_1(M)) \simeq \mathrm{Isom}(M)$, and is hence a finite group.

Proof. Let $\Gamma = \pi_1(M)$. We have a map

$$\theta : \mathrm{Isom}(M) \to \mathrm{Out}(\Gamma)$$

given by $f \mapsto [f_*]$, because f_* is an isomorphism of $\pi_1(M, x)$ onto $\pi_1(M, f(x))$ and basepoint change is given by conjugation, i.e, inner automorphisms. We show that the map θ is an isomorphism, and then conclude that the groups are finite by showing that $\mathrm{Isom}(M)$ contains finitely many homotopy classes.

Injectivity

Suppose $\theta(f) = [1]$. There is a lift $\widetilde{f} : \mathbb{H}^n \to \mathbb{H}^n$ of f such that $\widetilde{f} \circ \gamma = \gamma \circ \widetilde{f}$ for all $\gamma \in \Gamma$.

Let $\delta \neq 1$ be in the centralizer. As $\gamma \in \Gamma \setminus \{1\}$ is hyperbolic, with unique axis ℓ_γ, we get

$$\delta(\ell_\gamma) = \gamma(\delta(\ell_\gamma)) \implies \delta(\ell_\gamma) = \ell_\gamma,$$

and so δ is not parabolic. Let $F = \mathrm{Fix}(\delta)$. Then, for all $\gamma \in \Gamma \setminus \{1\}, \ell_\gamma \subset F$ and $\gamma(F) = F$.

Fix some $x_0 \in F$ and a line ℓ_0 through x_0 that is orthogonal to F. Then, for small ε,

$$\overline{N_\varepsilon(\ell_0)} \cap (\Gamma \setminus \{1\}) \cdot \overline{N_\varepsilon(\ell_0)} = \emptyset,$$

and so we get a closed subset of M that is not compact. $\rightarrow\leftarrow$

Surjectivity

Any automorphism of Γ is induced by a homotopy equivalence of M, because M is a $K(\Gamma, 1)$ space. Mostow rigidity gives an isometry f which induces the automorphism, and hence θ is surjective.

Finiteness

Since M is compact, there is $\rho > 0$ such that if $d(x, y) < \rho$ for $x, y \in M$, then the geodesic of length $d(x, y)$ joining x and y is unique. This means that for such x, y any convex combination is uniquely defined. Now, the group $\mathrm{Isom}(M)$ is compact with respect to the sup metric,

$$d(f_1, f_2) = \sup_{x \in M} d(f_1(x), f_2(x)).$$

But $d(f_1, f_2) < \rho$ implies they are homotopic, and there are finitely many such balls covering the group. Hence, $\mathrm{Isom}(M)$ contains only finitely many homotopy classes, thus proving that it is finite. $\square$

9 Further directions

We mention here briefly some extensions of Mostow's result. Firstly, Theorem (1) is due to Mostow only in dimension 3. Marden proved in 1974 a version of the result for hyperbolic $3-$manifolds with "cusps", which has now come to be Marden's isomorphism theorem and is a key result in the study of Kleinian groups that have parabolic elements. Prasad proved in 1974 that the result holds (for all dimensions ≥ 3) if we assume that the manifolds are of finite volume, weakening the assumption of compactness. Gromov's proof works almost identically in this setup, with one more idea: the thick-thin decomposition due to Margulis. Finally, Thurston and others proved the following result in the 1980's, that can be thought of as the ultimate version of Gromov's theorem (8).

Theorem 15

If $f : M_1 \to M_2$ is a smooth map such that $\mathrm{vol}(M_1) = |\deg f| \, \mathrm{vol}(M_2)$, then f is homotopic to a locally isometric covering of M_1 onto M_2, of degree $|\deg f|$.

A proof of this result is given in [Thu78].

References

[BP92] Riccardo Benedetti and Carlo Petronio. *Lectures on Hyperbolic geometry*. Springer, 1992.

[Mar07] Albert Marden. *Outer Circles*. Cambridge, 2007.

[Mar22] Bruno Martelli. An introduction to geometric topology, 2022.

[Thu78] William Thurston. The geometry and topology of 3-manifolds. *Lecture notes*, 1978.