Mostow Rigidity and Hyperbolic 3-Manifolds

a dissertation presented
by
Benjy Firester
to
The Department of Mathematics

in partial fulfillment of the requirements
for the degree of
Bachelor of Arts with Honors
in the subject of
Mathematics

Advised by
Curt McMullen

Harvard University
Cambridge, Massachusetts
May 2023

benjaminfirester@college.harvard.edu
917-887-6359

Mostow Rigidity and Hyperbolic 3-Manifolds

Abstract

Mostow rigidity is the remarkable theorem stating the uniqueness of hyperbolic structures on manifolds of dimension three or higher. Geometrically, this states that any homotopy equivalence between two hyperbolic manifolds can be uniquely deformed to an isometry; informally, in this context, geometry equals topology. Hyperbolic manifolds form an abundant and rich class of manifolds which provide further insight into other fields including topology, algebra, and dynamics. This rigidity means that any geometric invariant is also a topological invariant, and provides a powerful toolkit to understand these objects. We highlight the two and three dimensional cases in detail and show why rigidity only holds in dimensions three and higher. In dimension two, we exemplify the flexibility of hyperbolic structures encapsulated by a high dimensional moduli space that parameterizes the various hyperbolic structures.

The two main proofs of Mostow rigidity hinge on extending an initial map between two hyperbolic manifolds to the boundary of compactified hyperbolic space, the *sphere at infinity*. Mostow's original proof utilizes the ergodicity of the geodesic flow to show that this map is conformal, and therefore extends to a unique isometry on hyperbolic space, which further descends to an isometry between the original manifolds. Gromov's proof seeks to answer the question of how to directly capture the topological invariance of the hyperbolic volume. He defined a purely topological quantity which captures the complexity of the fundamental class of a manifold, and in the hyperbolic setting, computes the volume. The second main realization in this proof is that the fundamental group of a hyperbolic manifold algebraically encapsulates the geometric information of simplices of maximal volume, which are rigid above dimension two.

Contents

N.B. Throughout this thesis, the formal academic "we" is used. All the writing is my own, and all figures were made by me unless otherwise noted.

Acknowledgments

I WANT TO THANK, first and foremost, Curt McMullen, who beyond being an amazing advisor for this thesis, has been a wonderful mentor and mathematics teacher. I've been extremely fortunate to have learned from him over the past four years, a highlight of my undergraduate experience. Curt introduced me to the steep pitches of mathematics, and more importantly, how to approach mathematical problem-solving. When I am stuck on a problem, I find the next, right question to ask by thinking about what Curt would ask me if he were next to me. In my mathematical climb, Curt has held the ropes tightly and gone above and beyond to help me progress.

I am also deeply grateful for the teaching and mentorship of Joe Harris during my entire time at Harvard. Joe started me on my mathematical journey and has constantly supported me and pushed me further into the field. His hands-on geometric approach to math has shaped my own technique, and I continue to use his teachings every day. I am especially thankful for his support of my thesis work over the summer, helping me learn in a new field and pushing me to higher levels of understanding.

A tremendous thank you goes to Tristan Collins for taking a chance on me and taking me under his wing in my mathematics research. Tristan has been more than generous with his time and tutelage, supporting, teaching, and mentoring me, even remotely from across the world. Countless times, when I find a particularly creative solution in a math proof, I realize that I have used a "Tristan-ism" – finding the leaf that explains the tree – in my problem-solving approach. The mathematics research I have done under Tristan's mentorship over the past few years has been the most transformative part of my undergraduate journey in math. It was from Tristan that I began to learn how to do research in mathematics.

I owe my passion for differential geometry to Sébastien Picard, who catapulted me into this field through his love of the subject and charismatic teaching; and also to Valentino Tosatti, who introduced me to complex geometry and cultivated my love for the field. Their respective courses, Kähler Geometry and Non-Kähler Geometry, should have summed to a universal cover, but in fact left me yearning to learn more geometry.

Thank you to my friends who made Harvard my home, supported and pushed me, and always made me happy. Thank you to Raphael, my best friend, collaborator, and teacher, with whom time seemed to never end; and to Philip and Raluca, who made the Math Department lively and fun, the late night p-sets being some of my more cherished times; and to my Eliot House adoptive blockmates, Raphael, Isaac, Pedro, Thomas, David, and Arty, who gave me another home and some of my fondest memories. Thank you to Sílvia, the best friend I could have asked for, who brought infinite joy to even the busiest times; I certainly could not have done it without you.

Finally, thank you to my parents and siblings whom I love so much. Mom and Dad, you taught me my very first math and gave me every opportunity to pursue my goals, and thanks to you, I am living them now. Your support and love knows no bounds, and everything I have achieved is all because of you. Kalia and Ari, you never cease to show me love and support, and my admiration for you is unbounded; my brother Ari, for your immense generosity and compassion, and Kalia, for your creativity and passion. Additional thanks to Kalia for drawing some of the more difficult figures.

תודה לסבא וסבתא על האהבה והתמיכה והבית בו הפקתי את עבודתי החשובה ביותר.

1
Introduction

In this work, we present a robust exposition of Mostow rigidity; this remarkable result states that in dimensions at least three, hyperbolic structures are unique. We will focus on the fundamental differences between dimensions two and three, and the richness and abundance of hyperbolic 3-manifolds, through examples and computation. In higher dimensions, little is known about the geometric landscape of manifolds. These results will extend to higher dimensions and here we highlight the more accessible geometry present in dimension three, while the results are done in full generality. Throughout this exposition, we hope to explain and motivate the failure of rigidity proofs in dimension two and concretely demonstrate the flexibility of hyperbolic surfaces.

1.1 Mostow rigidity

Theorem 1.1.1 (Mostow rigidity)**.** *Let $f : M \to N$ be a homotopy equivalence between two finite volume hyperbolic manifolds of dimension $n \geq 3$. Then f is homotopic to an isometry.*

Geometry imposes an extremely detailed structure on a space, and with it comes attached a rich set of tools and invariants. Mostow rigidity therefore provides a bridge between results and computations in geometry to topology and algebra.

Corollary 1.1.2. *Any geometric invariant of a finite volume hyperbolic manifold M of dimension $n \geq 3$ is a topological invariant. Examples include:*

(i) the hyperbolic volume,

(ii) the length of the shortest geodesic, and

(iii) the spectrum of the Laplacian, notably the smallest eigenvalue.

Another corollary states that for hyperbolic manifolds, the weaker notion of homotopy equivalence implies the stronger notion of *homeomorphism*.

Corollary 1.1.3. *If M and N are homotopy equivalent hyperbolic manifolds of dimension at least three, they are homeomorphic.*

This result completely falls apart in similar settings such as spherical geometry. Algebraically, the structure of a hyperbolic manifold is entirely encoded in its fundamental group Γ, as such a manifold is always expressible as $\mathbb{H}^n/\Gamma$, a fact which we will go over in great detail in Chapter 2. This rigidity theorem allows us to understand algebraic facts about Γ.

Corollary 1.1.4. *The outer isomorphism group of Γ, namely $\mathrm{Aut}(\Gamma)/\mathrm{Inn}(\Gamma)$, is finite. It is canonically isomorphic to the isometries of $M = \mathbb{H}^n/\Gamma$.*

We can similarly deduce topological invariants from the algebra of Γ.

Corollary 1.1.5. *Any algebraic invariant of Γ is a topological invariant of $M = \mathbb{H}^n/\Gamma$. In dimension three, Γ is a discrete subgroup of $\mathrm{PSL}(2,\mathbb{C})$, and taking the trace of every element in Γ forms a number field which is an algebraic invariant associated to M as a topological space.*

1.1.1 Extending the map to the boundary sphere at infinity

The first step in both Mostow's and Gromov's proof of rigidity is to lift the initial map $f : M \to N$ not only to the universal cover $\tilde{f} : \mathbb{H}^n \to \mathbb{H}^n$, but to get an induced map on the sphere at infinity, which compactifies hyperbolic space radially. The main result here will be that $\tilde{f}$ is a *pseudo-isometry* meaning it almost preserves distances. The images of geodesics under $\tilde{f}$ will be *quasi-geodesics*. These quasi-geodesics lie within a bounded distance of a unique hyperbolic geodesic; notably, they will match up at the endpoints at infinity. This will mean that although a priori $\tilde{f}$ does not take geodesics to geodesics, it will indeed associate a unique geodesic and induces a pairing on geodesics. Geodesics are parameterized by pairs of distinct points on the sphere at infinity, so this will give us the induced map on the boundary spheres. The key property of hyperbolic geometry that will allow this is informally that the distance between two points p and q, both t away from a geodesic, grows exponentially, according to $\cosh(t)$ compared to their projections onto the geodesic. That is, the cost of traveling inefficiently in hyperbolic space grows exponentially.

1.1.2 Mostow's proof

In Chapter 3, we will detail Mostow's original proof, which utilizes the theory of quasiconformal mappings and ergodicity. Mostow proved that the map on the boundary is quasi-conformal, meaning it only distorts the local geometry in a bounded manner. Heuristically, infinitely small disks are mapped to ellipsoids of bounded eccentricity. This bound is called the *dilatation*, and when it is 1, these maps are called *conformal* and have extra structure giving regularity results. The conformal geometry at the sphere at infinity corresponds to the hyperbolic geometry in $\mathbb{H}^n$, and a key insight of Mostow is that in dimensions at least three, the induced map on the sphere at infinity is quasi-conformal. The proof finishes by using the ergodicity of the geodesic flow to improve the regularity of the map on the spheres at infinity to show that it is actually conformal, and therefore corresponds uniquely to an isometry on the interior to which we can homotopically deform f.

1.1.3 Gromov's proof

In Chapter 4, we present Gromov's proof, which immediately captures the relationship between topology and geometry in the hyperbolic setting. Having already known the theorem of Mostow rigidity, the first step of the Gromov proof answers the following question: How can we compute the hyperbolic volume, (which we know is a topological invariant), only using the topology? Gromov defines a purely topological quantity associated to the hyperbolic manifold using its fundamental class, called the *Gromov norm*, which in the hyperbolic setting is proportional to its hyperbolic volume.[1] Gromov's proof then proceeds by proving that the extended map on the boundary must preserve the maximal volume simplices (which are *regular ideal simplices*) that are maximally symmetric with vertices at infinity. This proof originally only worked in dimension three, where the maximal volume of tetrahedra was known to be maximally symmetric. It was extended in all higher dimensions by Haagerup and Munkholm [HM81] and we detail this result in Appendix A.

The second main insight of the Gromov proof is that the fundamental group Γ contains the geometric data of the regular ideal simplices. In dimension three, this corresponds to Γ understanding exactly which 4-tuples of points in $\mathbb{CP}^1$ have cross ratio $\zeta_6 = e^{\pi i/3}$. This is geometrically characterized by the induced map on the sphere at infinity which gives rise to a bijection on ideal regular simplices. In dimension two, all ideal triangles are the same, so this information is vacuous, which shows the lack of rigidity.

1.2 Hyperbolic 3-manifolds

In this report, we will rigorously define the necessary geometric, analytic, and algebraic structures used in hyperbolic geometry, with particular focus on the 2- and 3-dimensional cases where hyperbolic manifolds are plentiful. We include detailed constructions, hyperbolic computations, and figures. Intuitively, hyperbolic manifolds are characterized by having constant negative curvature. Locally, they look like a saddle curving along opposite directions as one rotates around a point. In dimensions two and three, *most* manifolds are hyperbolic, and in Chapter 2, we will give many motivating examples. We will detail the

[1] This should actually be called a semi-norm because it vanishes for many non-trivial topological objects including spheres.

various models of hyperbolic space in all dimensions and show how to translate between them, utilizing the best features of each model to perform explicit computations.

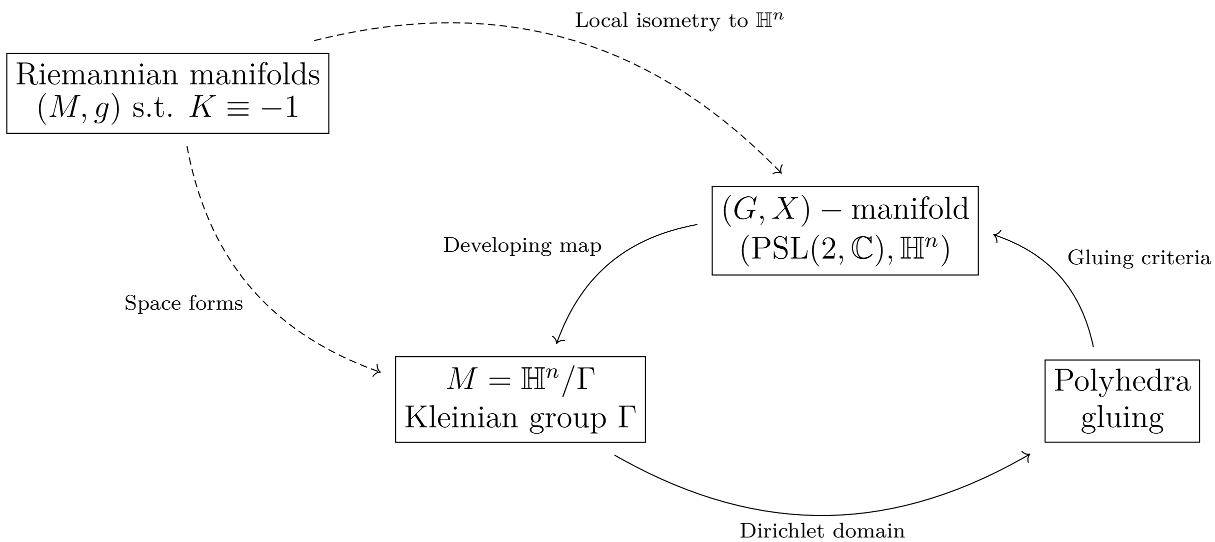

Figure 1.1: The models of complete hyperbolic manifolds and the constructions to translate among them.

There are three modern, equivalent viewpoints of hyperbolic manifolds, each with its own merits. To work fluently with this geometry, it is critical to be able to go back and forth between them. These three models construct M as a complete finite volume hyperbolic manifold:

(i) $M = \mathbb{H}^n/\Gamma$, a quotient of hyperbolic space by a discrete subgroup $\Gamma \subset \mathrm{Isom}(\mathbb{H}^n)$ of the isometries of hyperbolic space. Such a Γ is called *Kleinian*.

(ii) M is given a (G, X)-structure for $(G, X) = (\mathrm{Isom}(\mathbb{H}^n), \mathbb{H}^n)$. This means that locally we have charts on M that are modeled as $\mathbb{H}^n$ with transition functions that are isometries.

(iii) M is constructed by gluing hyperbolic polyhedra along pairwise isometric identifications of their boundary sides.

Translating among these models requires a few constructions and theorems that we flesh out in Chapter 2. The primary and most versatile definition is (i) which is detailed in Chapter 2 Section 2.3 after we define the requisite background on hyperbolic space and the isometry group. Starting from the local picture of M as a (G, X) manifold, we will define the *developing map* in Chapter 2 Section 2.4 which realizes M as $\mathbb{H}^n/\Gamma$. The idea here is to take the local picture of M as $\mathbb{H}^n$ given by the (G, X)-structure and *unfold* its universal cover along the isometries in $\mathbb{H}^n$. To realize $M = \mathbb{H}^n/\Gamma$ as the gluing of hyperbolic polyhedra, we must find a *fundamental domain* that models a polyhedron in $\mathbb{H}^n$ such that its faces can be glued together along isometries to create M. We define this construction in Chapter 2 Section 2.6.1, which is called a *Dirichlet domain* and can be made by taking a basepoint and finding its orbit under Γ in $\mathbb{H}^n$ and the fundamental domain is given as the points closest to this initial basepoint; the *hyperbolic Voronoi* region. Lastly, given a topological space expressed as a polyhedra gluing, to check that the topological space has a manifold structure, one must verify that it has local charts around the higher codimension faces to satisfy the local (G, X)-conditions. This is given by the Poincaré polyhedra theorem 2.6.4. In particular, in dimension three, the gluing conditions are extremely easy to verify and are equivalent to the vanishing of the Euler characteristic as detailed in Proposition 2.6.2.

There is a final definition coming from Riemannian geometry, and we will briefly establish its equivalence to the others in Chapter 2 Section 2.7. In the Riemannian setting, a hyperbolic manifold is given by a Riemannian metric (M, g) such that g has constant sectional curvature -1. We will show that when this is complete, its universal cover is $\mathbb{H}^n$, which shows this is equivalent to Definition (i) above. It can also be shown to be locally isometric to $\mathbb{H}^n$ showing equivalence to Definition (iii) above. Because there are not generally methods to produce Riemannian manifolds with constant sectional curvature, in the Riemannian setting, such structures are quite rare, so this perspective may be regarded as the least natural. See Figure 1.1 to diagram the constructions used to go among the models.

2

Hyperbolic geometry fundamentals and examples

Hyperbolic manifolds are characterized by constant negative sectional curvature. This local description means that any such manifold can be covered by charts diffeomorphic to $\mathbb{R}^n$ that carry metrics of constant negative curvature, motivating the definition of the hyperbolic metric on either the open ball or upper half-space $\mathbb{H}^n$. It is a remarkable fact that this local description can be made global using the developing map, giving rise to the classification of complete constant sectional curvature manifolds, also called *space forms*. This theorem states that any complete manifold of constant sectional curvature is a quotient of either the sphere S^n, flat Euclidean space $\mathbb{R}^n$, or the hyperbolic space $\mathbb{H}^n$. Therefore, a global description of hyperbolic manifolds is given as a quotient $M = \mathbb{H}^n/\Gamma$ of the hyperbolic plane by a discrete subgroup of its isometries. Such a group is called a *Kleinian group*, and it characterizes the entire geometry of a hyperbolic manifold. This group is naturally identified with the fundamental group $\pi_1(M)$. Utilizing the universal covering map $\pi : \mathbb{H}^n \to M$ of a hyperbolic manifold, we can understand the original manifold by examining its fundamental group; Mostow rigidity says that this is sufficient to uniquely determine the manifold in dimension at least three. Hyperbolic manifolds can be explicitly constructed using hyperbolic polyhedra, which act as puzzle pieces. When glued together properly, they assemble to form recognizable spaces and can geometrize many spaces that *a priori* are described only topologically.

Hyperbolic manifolds play the most important role in understanding the geometry of two- and three-dimensional manifolds, and likely in higher dimensions. In these low dimensional cases, there are *geometrization* theorems that state that all manifolds have natural geometric structures. In dimension two, this is given by the Uniformization theorem which states that any Riemann surface can be endowed with a metric of constant curvature, either flat, positive (spheres), or hyperbolic. All but finitely many of these will be hyperbolic. We detail this result in Section 2.9. In three dimensions, there are eight model geometries – the brilliant insight that is Thurston's geometrization conjecture, proven by Perelman using Ricci flow. This theorem states that any 3-manifold can be cut into pieces each of which has a geometric structure that aligns with one of just eight different types. Similarly to the dimension two case, the hyperbolic case occupies most of these manifolds. The idea that *most 3-manifolds are hyperbolic* can be made rigorous and demonstrates why this is the most important geometry to study (see Theorem 2.10.9). Little is known on the general structure of 4-manifolds or higher, although hyperbolic manifolds of all dimensions exist and are plentiful, and the geometric intuition and examples shown in dimension three here extend to higher dimensional cases. Hyperbolic manifolds further lie at the intersection of many fields of math and give insight into results in topology, algebra, and analysis. Their study is intimately related to physics, and the theory of special relativity can be succinctly expressed using the hyperbolic Minkowski metric.

In this chapter, we will give examples and pictures of hyperbolic manifolds of dimension two and three and explicitly demonstrate the fundamental differences between these two spaces of manifolds, and show why rigidity cannot hold in dimension two and give intuition as to where this comes from in dimension three. We will compute the area and volume of hyperbolic polyhedra in dimensions two and three and use them to compute volumes of various hyperbolic manifolds. Additionally, we will show the vastness of hyperbolic 3-manifolds and explain the *geometric topology* which gives a geometric structure on the space of all finite volume hyperbolic 3-manifolds and orders them based on volume.

The proof of Mostow rigidity requires some algebraic topology as well, and in an effort to produce a close to self-contained document, we detail the important theorems about hyperbolic manifolds in the topological setting. Because $\mathbb{H}^n$ is contractible, any hyperbolic manifold is an Eilenberg-Maclane space $K(\pi_1(M), 1)$, meaning that higher homotopy groups vanish. These spaces are the building blocks of topology. Mostow rigidity therefore provides a powerful bridge connecting geometry and topology; any geometric invariant is also a topological invariant, such as hyperbolic volume. In Section 2.12, we provide the necessary results to forge ahead with Mostow rigidity and more details of their proofs are given in Appendix B.

2.1 Models of hyperbolic space

Definition 2.1.1 (Upper half-space model)**.** The primary model of hyperbolic space is the upper half-space, denoted $\mathbb{H}^n$. The space $\mathbb{H}^n = \{(x^1, \ldots, x^n) \in \mathbb{R}^n : x^n > 0\}$ is endowed with the metric $g_{ij} = \frac{\delta_{ij}}{(x^n)^2}$. Since this norm is always positive, as the point 0 is not in the upper half-space, this metric is conformal to the Euclidean metric (see below conformal metrics 2.1.9).

In the upper half-space model, we have a natural boundary where $x^n = 0$. However, it is often helpful to also consider a compactification of this boundary at infinity. To do that, we can conformally map the upper half-space into the interior of a disk.

Definition 2.1.2 (Open ball model)**.** The space $\mathbb{B}^n = \{(x^1, \ldots, x^n) \in \mathbb{R}^n : \|x\| < 1\}$ is the open unit ball, and we give it the metric $g_{ij} = \frac{4}{(1-\|x\|^2)^2}\delta_{ij}$, a conformal scaling of the Euclidean metric.

Another natural way to realize hyperbolic space is to ask if it can be isometrically embedded as a hypersurface. To achieve this, we take the so-called *sphere of radius i* in $\mathbb{R}^{n,1}$, the $(n+1)$-dimensional space endowed with a quadratic form of signature $(n, 1)$.

Definition 2.1.3 (Minkowski model)**.** We seek to find a well-curved copy of $\mathbb{R}^n$ sitting inside $\mathbb{R}^{n+1}$ such that the induced metric is the hyperbolic metric. To do this, let x be a coordinate on $\mathbb{R}^n$ and we add a variable t. We define

$$\mathcal{H}^n = \{(x, t) \in \mathbb{R}^n \times \mathbb{R} : |x|^2 - t^2 = -1, t > 0\}$$

to be the 1-sheeted hyperboloid, or the so-called sphere of radius i for a quadratic form of signature $(n, 1)$ called the Minkowski metric, which we may denote $\langle -, - \rangle_{n,1}$.

The condition of $t > 0$ takes a single sheet of a real two-sheeted hyperboloid. Rearranging the equation, we can see that $|x|^2 = t^2 - 1$, which can only be solved when $t \geq 1$. Therefore, we can take the plane $t = 1$, and this plane meets the Minkowski model tangentially at the point where $x = 0$ and $t = 1$. The entire Minkowski plane lies inside the *light cone* where $|x|^2 = t^2$, which when intersected with the plane at $t = 1$ bounds a unit disk centered at $x = 0$. We can project the Minkowski model into this disk by taking the unique line through any point and the origin, and taking its intersection with the above disk.

Definition 2.1.4 (Klein model)**.** We can projectivize the Minkowski model and get that $\mathcal{H}^n$ will lie in $\mathbb{RP}^n$ to achieve the Klein model. Under this map, $\mathcal{H}^n$ will become the unit ball in the projective coordinates $[x : 1]$. We may denote this model as $\mathcal{K}^n$.

2.1.1 Geodesics

Geodesics are the geometric notion of a straight line for Riemannian manifolds. Given a Riemannian manifold (M, g), the metric is given as g which is a family of smoothly varying inner products of tangent vectors at every point in M. Given a smooth curve $\gamma : [0, 1] \to M$, we can define the *energy functional* as

$$\int_0^1 |\dot{\gamma}(t)|^2 \, dt$$

where we measure the length using g. Motivated by physics, geodesics are critical points of the energy functional, or equivalently, solve the Euler-Lagrange type differential equation:

$$\ddot{\gamma}^k + \Gamma^k_{ij}\dot{\gamma}^i\dot{\gamma}^j = 0 \tag{2.1}$$

where the repeated indices in a product mean that we sum over all components i and j, and this is true for all k. The symbols Γ_{ij}^k are the *Christoffel symbols* of g defined in local coordinates[1] as

$$\Gamma_{ij}^k = \frac{1}{2} g^{k\ell}(\partial_i g_{\ell j} + \partial_j g_{i\ell} - \partial_\ell g_{ij}). \tag{2.2}$$

This equation is a second-order ordinary differential equation and has a short-time unique solution given by the Picard-Lindelöf theorem. Geodesics have constant speed, so to consider the space of geodesics, we will normalize this speed to 1 to avoid multiple geodesics tracing out the same path, only differing in velocity, (constructed below in Remark 2.1.8 after the characterization of hyperbolic geodesics).

To get a handle on hyperbolic space, we will want to have familiarity with what the geometry looks like, so we must understand the geodesics and the angles among them. The various models have different advantages in computing and analyzing geodesics. In $\mathbb{H}^n$ and $\mathbb{B}^n$, the angles between geodesics will agree with the Euclidean angle between their tangent lines at the point of intersection. In $\mathcal{H}^n$, the geodesics can be computed with linear algebra. In the Klein model, the geodesics are straight lines, but the angles and speeds along them vary and disagree with the Euclidean parameterizations. Table 2.1 below details the geodesics in each model:

Model	Geodesics	Properties
$\mathbb{H}^n$	Vertical lines and semicircles orthogonal to $\{x^n = 0\}$ with centers on the plane $\{x^n = 0\}$	Angles agree with Euclidean angles, boundary is $\{x^n = 0\} \cup \infty$
$\mathbb{B}^n$	Circular arcs that meet the boundary sphere at right angles	Angles agree with Euclidean angles, compactified $\overline{\mathbb{B}^n}$ is Euclidean closure
$\mathcal{H}^n$	Intersections of $\mathcal{H}^n$ with planes through the origin of $\mathbb{R}^{n,1}$	Geodesics are parameterized by hyperbolic trigonometric functions cosh and sinh
$\mathcal{K}^n$	Straight lines	Convexity is the same as Euclidean convexity, but angles do not agree

Table 2.1: Geodesics in each model and properties.

In both the upper half-space model and the Poinaré ball model, the geodesics are given as arcs of a circle that are perpendicular to the boundary. Geodesics in the Minkowski and Klein model are given as intersections with two dimensional linear subspaces in $\mathbb{R}^{n,1}$. In the upper half-space and Poincaré disk models, the geodesics do not align with straight lines in the flat setting. The benefit is that the hyperbolic angles in these models align

[1]A manifold has local coordinates given by a smooth atlas. An atlas is a collection of open subsets $\{U_i\}$ of M that cover M, meaning $\bigcup_i U_i = M$, such that each open subset is diffeomorphic to an open subset of via maps $\varphi_i : U_i \to V_i \subset \mathbb{R}^n$. On the overlaps $U_i \cap U_j$, the map $\varphi_i \circ \varphi_j^{-1}$ is a smooth function from $V_j \to V_i$ in the standard sense. The V_i have standard coordinates in $\mathbb{R}^n$ which we say are local coordinates on M.

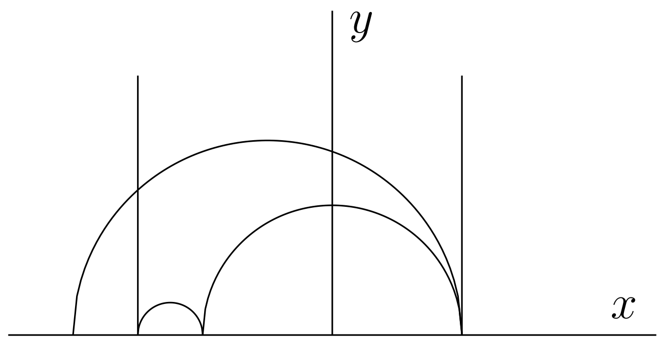

Figure 2.1: Geodesics in $\mathbb{H}$.

with the flat angles. In the Klein model, the geodesics will be straight lines, but the hyperbolic angles will be distorted and the length of the curve does not agree with the Euclidean length.

These models come in natural pairs as can be seen in Table 2.1. The upper half-space and Poincaré ball are conformally related to the standard Euclidean, so analysis on one carries over to the other (Subsection 2.1.2 will detail more discussion on conformal metrics). The relationship between $\mathcal{H}^n$ and $\mathcal{K}^n$ is given by projectivization, which defines the structure on $\mathcal{K}^n$ to be induced by this map as an isometry that is also an isomorphism. Therefore, we can study just upper half-space $\mathbb{H}^n$ and $\mathcal{H}^n$ to understand all geodesics. Furthermore, the rotational symmetries of these spaces – rotation about the x^n-axis in $\mathbb{H}^n$, and rotation about the t-axis in $\mathcal{H}^n$ – reduces all the analysis to studying simply the two-dimensional case.

Example 2.1.5 (Upper-half plane geodesics). Let $\mathbb{H} = \{z \in \mathbb{C} : \mathrm{Im}(z) > 0\}$ with metric $g_{ij} = \frac{1}{y^2}\delta_{ij}$. One model geodesic is the imaginary axis. The group of $\mathrm{PSL}(2, \mathbb{R})$ can carry this geodesic to any vertical line or semicircle perpendicular to the real line, and below we will show that this group in fact acts by isometries. Any two points are either with the same real part and can be connected by a vertical line, or are on the semicircle perpendicular to the real line centered at the unique real number equidistant from both points. See Figure 2.1. Therefore, any two points are connected by the image of this geodesic under some isometry, and since geodesics between two points are unique, this demonstrates all of them. Notably, $\mathbb{H}$ is geodesically complete and from the Hopf-Rinow theorem this tells us that the exponential map is defined everywhere (see Section 2.7). There are two ways to show that the vertical imaginary line is a geodesic. Firstly, we can argue that geodesics are length minimizing curves. Indeed we can compute the length $\gamma(t) : [a, b] \to \mathbb{H}$ maps

$\gamma : t \mapsto (0, t)$. Let ℓ be any piecewise C^1 arc adjoining $\gamma(a)$ to $\gamma(b)$.

$$\begin{aligned} L(\ell) &= \int_a^b \left|\frac{d\ell}{dt}\right| dt \\ &= \int_a^b \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}\frac{dt}{y} \\ &\geq \int_a^b \left|\frac{dy}{dt}\right|\frac{dt}{y} \\ &\geq \int_a^b \frac{dy}{y} \\ &= L(\gamma) \end{aligned} \tag{2.3}$$

thereby demonstrating that γ is a geodesic.

The geodesics can also be directly computed using the geodesic differential equation from above

$$\ddot{\gamma}^k + \Gamma^k_{ij}\dot{\gamma}^i\dot{\gamma}^j = 0 \tag{2.4}$$

for $\dot{\gamma}$ representing $\frac{d\gamma}{dt}$, the tangent vector to γ at time t. The Christoffel symbols for $\mathbb{H}^2$ are

$$\Gamma^x_{ij} = \begin{pmatrix} 0 & -\frac{1}{y} \\ -\frac{1}{y} & 0 \end{pmatrix}, \qquad \Gamma^y_{ij} = \begin{pmatrix} \frac{1}{y} & 0 \\ 0 & -\frac{1}{y} \end{pmatrix} \tag{2.5}$$

so the geodesic equations 2.4 using the Christoffel symbols 2.5 becomes

$$\ddot{x} - \frac{2}{y}\dot{x}\dot{y} = 0, \qquad \ddot{y} + \frac{1}{y}\dot{x}^2 - \frac{1}{y}\dot{y}^2 = 0 \tag{2.6}$$

to which we recognize that the imaginary axis of $x = 0$ solves the first equation and reduces the second one to $y\ddot{y} = \dot{y}^2$ solved by $y = e^t$.

To demonstrate that the isometry group is $\mathrm{PSL}(2, \mathbb{R})$, the action is given by

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}(2, \mathbb{R})$$

on $z = x + iy$ by the linear fractional transformation $Az = \dfrac{az + b}{cz + d}$ which is invariant under the action of $\pm I$. Let x', y' be the new coordinates under the action of A on x, y. Explicitly, this is written out as

$$\frac{ax + b + ayi}{cx + d + cyi}$$

Killing the imaginary part of the denominator gives this as:

$$\frac{(ad - bc)y}{(cx + d)^2 + (cy)^2} > 0$$

which is positive since $\det A = 1$. This whole part is:

$$\frac{ac(x^2+y^2)+(ad+bc)x+bd}{(cx+d)^2+(cy)^2}+\frac{y}{(cx+d)^2+(cy)^2}i$$

Differentiating both of these with respect to t and plugging into:

$$\frac{\dot{x}'^2+\dot{y}'^2}{y'(t)^2}$$

This gives the expression of the metric in the new coordinates as

$$dx'^2=\frac{\left(x'(t)\left(c^2x(t)^2-c^2y(t)^2+2cdx(t)+d^2\right)+2cy(t)y'(t)(cx(t)+d)\right)^2}{\left(c^2x(t)^2+c^2y(t)^2+2cdx(t)+d^2\right)^4}$$

and

$$dy'^2=\frac{\left(c^2y(t)^2y'(t)+2cy(t)x'(t)(cx(t)+d)-y'(t)(cx(t)+d)^2\right)^2}{\left(c^2x(t)^2+c^2y(t)^2+2cdx(t)+d^2\right)^4}.$$

Simplifying this gives

$$\frac{\dot{x}'^2+\dot{y}'^2}{y'(t)^2}=\frac{\dot{x}^2+\dot{y}^2}{y(t)^2},$$

so the action of A gives an isometry.

The theory of linear fractional transformations can be used to show that the open unit circle $\mathbb{C}\supset\Delta=\{z:|z|<1\}$ carries a hyperbolic metric given by the pullback of the hyperbolic metric on the upper half plane along the linear fractional transformation carrying the disk to the upper half plane such as $\begin{pmatrix}1 & -i\\ 1 & i\end{pmatrix}$. This discussion of the geometry of hyperbolic space will be useful in the uniformization theorem.

This extends to higher dimensions and gives a conformal isometry between $\mathbb{H}^n\to\mathbb{B}^n$, so we know it must preserve angles and take circles to circles, so the above result further proves that the geodesics in $\mathbb{B}^n$ are circular arcs orthogonal to the boundary as claimed in Table 2.1. An explicit conformal isometry from $\mathbb{H}^n\to\mathbb{B}^n$ is given by mapping the south pole $s=(0,\dots,0,-1)$ to infinity using the map $p:\mathbb{B}^n\to\mathbb{H}^n$ defined by $p:x\mapsto s+\frac{2(x-s)}{\|x-s\|^2}$.

Example 2.1.6 (Geodesics in $\mathcal{H}^n$). Let $x\in\mathcal{H}^n$ be a point in the Minkowski model and let $y\in T_x\mathcal{H}^n$ be a point such that $\langle y,y\rangle_{n,1}=1$ where $\langle -,-\rangle_{n,1}$ is the Minkowski metric of signature $(n,1)$ defined as $|x|^2-t^2$ above. The geodesic starting from x with initial velocity y is given as

$$\gamma(t)=\cosh(t)\,x+\sinh(t)\,y \tag{2.7}$$

notably, this is the intersection of $\mathcal{H}^n$ with a linear plane of dimension 2 going through the origin.

To see this result, let W be the plane generated by x and y. We define γ to be the maximally extended geodesic starting at x with velocity y. Let $\phi\in\mathrm{O}(n,1)$ be an orthogonal matrix (with respect to the Minkowski inner product) such that ϕ acts on W by the identity and acts on the orthogonal complement to W by negation. Therefore, $\phi(x) = x$ and $d_x\phi(y) = y$ implying that γ is invariant under ϕ and therefore contained in $W \cap \mathcal{H}^n$. We now realize that the mapping defined above in Equation 2.7 is of unit length and parameterizes $W \cap \mathcal{H}^n$ establishing the desired result.

Notably, projecting the linear planes to the Klein model $\mathcal{K}^n$ will produce straight lines as this will be the intersection of a plane through the origin, and the plane $\{t = 1\}$ (restricted to the unit disk), and the intersection of two planes will be a line. This works because the projection map from $\mathcal{H}^n \to \mathcal{K}^n$ is a global isometry by construction.

Example 2.1.7. The distance from a point $z = x + iy$ to the imaginary axis in $\mathbb{H}$ can be computed using the inverse of the hyperbolic sine function as $\sinh^{-1}(|x|/y)$. We need the unique geodesic that passes through the imaginary axis at a perpendicular angle and goes through the point $x + iy$. This means that it must pass through $(|x|^2 + |y|^2)i$ on the imaginary axis as the center of the circle must be the origin. We can use the isometry

$$\begin{pmatrix} r & 0 \\ 0 & r^{-1} \end{pmatrix}$$

for $r > 0$ real which acts on z by $\frac{rz+0}{0z+r^{-1}} = r^2 z$. This takes the imaginary axis to itself, so therefore we can reduce this problem by taking this isometry and assume that $|x|^2 + |y|^2 = 1$, so we are finding the distance to the point i using the above observation. We reflect to assume that $x > 0$.

We take the curve $\exp(i\theta)$ which is the geodesic that passes through both z and i. We need to take θ going from $\tan^{-1}(y/x)$ to $\frac{\pi}{2}$. The parameterization is $x = \cos(\theta)$ and $y = \sin(\theta)$, so the arc-length form is $\frac{1}{y}$. In terms of θ, we can compute $y = \sin\theta$. Therefore, this length is computed as

$$\int_{\tan^{-1}(y/x)}^{\frac{\pi}{2}} \frac{1}{\sin\theta}\, d\theta.$$

We can compute $\int \csc(\theta)\, d\theta = -\log|\csc(\theta) + \cot(\theta)| + C$. At $\theta = \pi/2$, this is $-\log(1) = 0$. Therefore, this length is

$$\log|\csc(\tan^{-1}(y/x)) + \cot(\tan^{-1}(y/x))| = \log|(1+x)/y|.$$

We have an exponential definition of the hyperbolic trigonometric sine function as

$$\sinh(x) = \frac{e^x - e^{-x}}{2} \implies \sinh^{-1}(x) = \log|x + \sqrt{x^2+1}|.$$

Recognizing this term above, we can rearrange the previous computation to a more compact form as

$$\begin{aligned} \log|(1+x)/y| &= \log|1/y + x/y| \\ &= \log|x/y + \sqrt{1/y^2}| \end{aligned}$$

$$= \log |x/y + \sqrt{\frac{x^2 + y^2}{y^2}}|$$
$$= \log |x/y + \sqrt{(x/y)^2 + 1}|$$
$$= \sinh^{-1}(x/y)$$

demonstrating the desired result.

Remark 2.1.8 (Space of geodesics). Most easily seen in $\mathbb{B}^n$ and $\mathcal{K}^n$, any geodesic can be infinitely extended to meet two distinct points on the boundary sphere at infinity S^{n-1}_∞. Therefore, by normalizing to unit speed, we can topologize the space of oriented geodesics as $S^{n-1}_\infty \times S^{n-1}_\infty \setminus \Delta$ where Δ is the diagonal subset $\Delta = \{(x, x) \in S^{n-1}_\infty \times S^{n-1}_\infty\}$. We impose that this is the space of oriented geodesics by imposing the direction of the geodesic given by $(x, y) \in S^{n-1}_\infty \times S^{n-1}_\infty$ as traveling from x towards y at unit speed.

2.1.2 Conformal metrics

Definition 2.1.9 (Conformal metrics)**.** Let g_{ij} be a metric on $\mathbb{R}^n$ (or the local presentation of a metric on a manifold). A metric h is said to be conformal to g if $h = e^f g$ is a positive rescaling of g at every point.

Proposition 2.1.10 (Curvature of conformal metrics)**.** *Let $g_{ij} = F^{-2}\delta_{ij}$ for a nowhere vanishing function F on some subset $\Omega \subset \mathbb{R}^n$ be a conformal scaling of the Euclidean metric. For $f = \log F$, the sectional curvature K_{ij} is given by*

$$K_{ij} = (\partial_i^2 f + \partial_j^2 f + (\partial_i f)^2 + (\partial_j f)^2 - \sum_\ell (\partial_\ell f)^2)F^2. \tag{2.8}$$

Proof. The inverse metric is given by $g^{ij} = F^2\delta^{ij}$ which we use to compute the Christoffel symbols defined in Equation 2.2. In the case of the conformal metric in, the computation is given as

$$\Gamma^k_{ij} = \frac{1}{2}F^2(\partial_i F^{-2}\delta_{\ell j} + \partial_j F^{-2}\delta_{i\ell} - \partial_\ell F^{-2}\delta_{ij}) = -\partial_i f\delta_{jk} - \partial_j f\delta_{ik} + \partial_k f\delta_{ij} \tag{2.9}$$

using the following expression of derivatives of the metric

$$\partial_k g_{ij} = \partial_k F^2\delta_{ij} = -\frac{2}{F^3}\partial_k F = -\frac{2}{F^2}\frac{\partial_k F}{F} = -\frac{2}{F^2}\partial_k(\log F) = -\frac{2}{F^2}\partial_k f.$$

so the prefactor term $\frac{1}{2}F^2$ cancels leaving only derivatives of f thereby proving the last equality in Equation 2.9. Notably, since each term has a δ_{ij} in it, if all three indices are distinct, this kills each term and the Christoffel symbol vanishes.

The expression of the Riemann curvature defined as

$$R^i{}_{jk\ell} = \partial_k\Gamma^i_{\ell j} - \partial_\ell\Gamma^i_{kj} + \Gamma^i_{kp}\Gamma^p_{\ell j} - \Gamma^i_{\ell p}\Gamma^p_{kj}$$

can now be simplified using the standard symmetries and the previous observation about the vanishing of the Christoffel symbols. If all four indices are distinct, the previous observation would show that the second half vanishes leaving just the first terms. Furthermore, these terms also vanish since all the derivatives of the Christoffel symbols with distinct indices vanish as well as seen by their formula in Equation 2.9. Therefore, at least two indices must be the same, so we can compute all the non-zero Christoffel symbols

$$\Gamma^i_{ij} = \Gamma^i_{ji} = -\partial_i f, \quad \Gamma^j_{ii} = \partial_j f, \quad \Gamma^i_{ii} = -\partial_i f \tag{2.10}$$

and the term of the Riemann tensor we will need is

$$R_{ijij} = R^k{}_{iji} g_{jk} = F^{-2} R^j{}_{iji} \tag{2.11}$$

to which Equation 2.10 can be applied to compute

$$F^2 R_{ijij} = \partial_j \Gamma^j_{ii} - \partial_i \Gamma^j_{ji} + \Gamma^k_{ii}\Gamma^j_{jk} - \Gamma^k_{ji}\Gamma^j_{ik} = \partial_j^2 f + \partial_i^2 f + (\partial_i f)^2 + (\partial_j f)^2 - \sum_k (\partial_k f)^2. \tag{2.12}$$

This can now compute directly the sectional curvature $K_{ij} = \frac{R_{ijij}}{g_{ii}g_{jj} - g_{ij}^2}$ which in this setting is

$$K_{ij} = F^4 R_{ijij} = F^2(\partial_j^2 f + \partial_i^2 f + (\partial_i f)^2 + (\partial_j f)^2 - \sum_k (\partial_k f)^2) \tag{2.13}$$

which is Equation 2.8 as desired. □

We can now specialize to where F is defined as in definitions 2.1.1 and 2.1.2 to compute that indeed the sectional curvature of the hyperbolic models is -1. For the upper half-space, this reduces considerably since $F = x^n$ so most of its derivatives are 0. In this case, we can use Equation 2.13 to compute K_{ij} noting that any derivative must be n to not vanish, so if neither i nor j are n, then this reduces to the single term in the sum where $k = n$ and we get

$$K_{ij} = (x^n)^2(-\partial_n \log x^n)^2 = -(x^n)^2 \frac{1}{(x^n)^2} = -1. \tag{2.14}$$

Now suppose that $i = n$ and $j \neq n$ (and by symmetry of the Riemann curvature tensor, this is also when $i \neq n$ and $j = n$), we compute

$$K_{nj} = (x^n)^2(-\partial_n^2 \log x^n + (\partial_n \log x^n)^2 - (\partial_n \log x^n)^2) = (x^n)^2 \left(\frac{1}{(x^n)^2} + \frac{1}{(x^n)^2} - \frac{1}{(x^n)^2}\right) = -1. \tag{2.15}$$

Finally, for $i = j = n$ the computation similarly gives

$$K_{nn} = (x^n)^2(2\,\partial_n^2 \log x^n + 2\,(\partial_n \log x^n)) = (x^n)^2 \left(-\frac{2}{(x^n)^2} + \frac{2}{(x^n)^2} - \frac{1}{(x^n)^2}\right) = -1. \tag{2.16}$$

Equation 2.13 similarly verifies that $K_{ij} = 1$ for the ball metric in the ball model from definition 2.1.2, however the conformal factor of $F = 2(1 - \|x\|^2)^2$ has derivatives in each variable, so it becomes much more complex. Instead, this can be verified by exhibiting an isometry between these spaces. We gave the example of this map by sending the south pole $s = (0, \ldots, 0, -1)$ to infinity using the map $p : \mathbb{B}^n \to \mathbb{H}^n$ defined by $p : x \mapsto s + \frac{2(x-s)}{\|x-s\|^2}$.

2.2 Completeness of hyperbolic space

The most important feature of the hyperbolic metric is that it is *complete*. Completeness has a few different definitions, but an important consequence of this is that any two points can be joined by a unique geodesic of minimal length.

Definition 2.2.1 (Geodesic completeness)**.** A Riemannian manifolds (M, g) is said to be *geodesically complete* if every geodesic can be extended infinitely.

As stated above, geodesics always have a short time existence. However, these short-time solutions do not always extend to infinite time solutions. Consider the Euclidean plane with the origin removed. We can start a path along the x-axis from the point $(1, 0)$ along the geodesic, however, it's length can never exceed one because the origin is not in this space. This is an example of a manifold that is not geodesically complete, nor is it metrically complete.[2] Notably, the distance between $(1, 0)$ and $(-1, 0)$ is two as defined as the infimum of lengths of all curves between those points, but this is never achieved by a geodesic.

Theorem 2.2.2 (Hopf-Rinow)**.** *A Riemannian manifold is metrically complete if and only if it is geodesically complete.*

For a proof of this theorem, we direct the reader to do Carmo [Car92] or other introductory textbooks on Riemannian geometry. Knowing this, we can now say a space is *complete* unambiguously. In Remark 2.1.8, we used the geometry of the geodesics in either $\mathbb{B}^n$ or $\mathcal{K}^n$ to note that any geodesic can be extended to infinity. In $\mathcal{K}^n$, it is further seen that any two points are connected by a unique geodesic, namely by using the Euclidean straight line. We formalize this intuition in the following proposition.

Proposition 2.2.3. $\mathbb{H}^n$ *is complete.*

Proof. $\mathbb{H}^n$ is seen to be complete by studying its geodesics and verifying that any two points can be connected by a unique geodesic. To examine the geodesics, it is sufficient to study the isometries of the upper-half of $\mathbb{R}^n$ that do not involve x^n as these will also be isometries if $\mathbb{H}^n$. In the $\mathbb{R}^{n-1}$ plane excluding x^n, the isometries agree, so we can study only the isometries in the plane spanned by x^1 and x^n. This is now the standard upper-half plane which is $\{z \in \mathbb{C} : \operatorname{Im}(z) > 0\}$ with metric $g_{ij} = \frac{1}{y^2}\delta_{ij}$. Any two points in $\mathbb{H}^n$

[2] A metric space X is (metrically) complete if we every Cauchy sequence has a limit. A Cauchy sequence is a sequence of points $x_n \in X$ such that for any $\varepsilon > 0$ there exists an N such that for all x_i, x_j where $i, j > N$, the distance $d(x_i, x_j) < \varepsilon$.

can be considered in $\mathbb{H}$ by taking the plane defined by them subject to the constraint that its projection along x^n is a line, that is perpendicular to the $x^n = 0$ plane. This justifies the reduction to studying just the upper-half plane $\mathbb{H}$ as detailed in Example 2.1.5. We finally appeal to the Hopf-Rinow theorem that says that completeness is equivalent to any geodesic having infinite time extension, which is the case as geometrically exemplified in Example 2.1.5 and Figure 2.1. $\square$

2.3 Isometries of $\mathbb{H}^n$

Isometries are maps that preserve distances, i.e. $f : X \to Y$ is an isometry if

$$\forall\, x, x' \in X, \quad d_X(x, x') = d_Y(f(x), f(x')).$$

Notably, isometries take geodesics to geodesics. The isometries of hyperbolic space can be represented as matrix groups, in particular, they are Lie groups[3], and their structure can be categorized by properties of their matrix representatives. We classify only the orientation preserving isometries as any orientation reversing isometry can be computed in the upper-half model by precomposition with the simple map $x_1 \mapsto -x_1$. In three dimensions, we consider the sphere at infinity to be the Riemann sphere $\mathbb{C} \cup \{\infty\}$. The hyperbolic geometry of $\mathbb{H}^3$ on the interior extends to conformal geometry on the boundary S^2. This is a key realization used in the Mostow proof to find conformal maps on the boundary that correspond to isometries in the interior. Therefore, the isometry group is characterized by $\mathrm{PSL}(2, \mathbb{C})$. This more generally extends and can be understood because of the completeness of hyperbolic space. Any isometry can be characterized by how it maps geodesics to geodesics, which correspond to pairs of points on the sphere at infinity. Understanding isometries corresponds to understanding conformal maps on the boundary space. We make this rigorous in the following theorem.

Theorem 2.3.1. *Isometries of $\mathbb{H}^n$ for $n \geq 2$ correspond with conformal automorphisms of the boundary sphere at infinity S^{n-1}_∞. Notably, we have that $\mathrm{Isom}(\mathbb{H}^2) = \mathrm{PSL}(2, \mathbb{R})$ and $\mathrm{Isom}(\mathbb{H}^3) = \mathrm{PSL}(2, \mathbb{C})$.*

In fact, we showed the identification $\mathrm{Isom}(\mathbb{H}^2) = \mathrm{PSL}(2, \mathbb{R})$ in the computations of the geodesics of the upper half-plane in Example 2.1.5.

Proof. We examine the geometry of isometries on $\mathbb{H}^n$ and how the extend to the boundary. Firstly, a reflection through any hyperplane H^{n-1} extends to a a reflection of S^{n-1}_∞ about a "circumference" sphere S^{n-2} which is a conformal map. The converse of the above construction follows as well. Take any given S^{n-2} and it bounds a hyperbolic hyperplane H^{n-1} and the reflection through this sphere can be extended on the interior to a hyperbolic reflection about H^{n-1}.

[3]Lie groups are smooth manifolds endowed with a group structure where the multiplication and inversion maps are smooth.

This completes the proof as both the isometry group of $\mathbb{H}^n$ and the conformal automorphisms of sphere are generated by their respective reflections. Therefore, any isometry extends to a conformal automorphism of the boundary, and conversely, a conformal automorphism of the sphere at infinity can be filled in to an isometry.

To verify that this is a one-to-one correspondence, consider an isometry that induces the identity automorphism on the boundary sphere at infinity. By characterizing all geodesics by their unique endpoints, we realize that this leaves every geodesic invariant. It must therefore be the identity. □

In two and three dimensions, the eigenvalues of matrices in $\mathrm{PSL}(2,\mathbb{R})$ and $\mathrm{PSL}(2,\mathbb{C})$ can be used to understand the different types of isometries. There are four categories of orientation preserving isometries that are easiest to describe across the various models of hyperbolic space. The fourth description is unique to $\mathrm{PSL}(2,\mathbb{C})$, but the first three work in either two or three dimensions.

(i) *Elliptic*: An isometry is called elliptic if it can be described in the Poincaré disk model as pure rotations across a geodesic which is a (Euclidean) straight line through the origin. These fix two points in the sphere at infinity. The trace of the matrix representing an elliptic isometry is real and contained in $(-2,2)$.

(ii) *Parabolic*: An isometry is called parabolic if it can be expressed in the upper half-space model as a translation. These isometries fix only a single point in the sphere at infinity, namely ∞ in the above description. The trace of the matrix representing a parabolic isometry is real and either ± 2.

(iii) *Hyperbolic*: An isometry is called hyperbolic if in the ball model, they can be realized as "translation" in the direction of a diameter. Pick a (Euclidean) straight geodesic through the origin and a hyperbolic isometry preserves the endpoints and shifts points towards the one endpoint of this diameter. The trace of the matrix representing a hyperbolic isometry is real and is greater than 2 in norm.

(iv) *Loxodromic*: An isometry is called loxodromic if it is hyperbolic and also rotates around the the diameter geodesic described above. The trace of the matrix representing a loxodromic isometry is not real.

2.3.1 $K-A-N$ subgroups and decomposition

For $\mathbb{H}^2$ and $\mathbb{H}^3$, we have explicit representations of the isometry groups of the hyperbolic plane and hyperbolic space as $\mathrm{PSL}(2,\mathbb{R})$ and $\mathrm{PSL}(2,\mathbb{C})$ respectively. We can analyze these as Lie groups and give some important properties that we will use in the proofs of Mostow rigidity. The main result we will need is the property of *unimodularity*, which means there exists a left- and right-invariant Haar measure. In both cases, we can express the homogeneous space of $\mathbb{H}^n$ for $n = 2,3$ as $\mathrm{SL}(2,\mathbb{R})/\mathrm{SO}(2)$ or $\mathrm{SL}(2,\mathbb{C})/(2)$ and the difference between these is really only the underlying field of $\mathbb{R}$ compared to $\mathbb{C}$, so they share a lot of desirable properties. Let $G = \mathrm{PSL}(2,k)$ for $k=\mathbb{R}$ or $k=\mathbb{C}$ for $n=2$ or $n=3$ respectively.

Notably, they are even of the same dimension when considered over their respective fields $\mathbb{R}$ and $\mathbb{C}$. We call the subgroups $K = \mathrm{SO}(2)$ or $K = \mathrm{SU}(2)$ for $n = 2$ or 3 respectively. We also define subgroups

$$A = \left\{ a_r = \begin{pmatrix} e^{r/2} & 0 \\ 0 & e^{-r/2} \end{pmatrix} \right\}, \qquad A \supset A^+ = \{a_r : r \geq 0\}$$

which make sense in both $n = 2$ and $n = 3$. The last subgroup is given as

$$N = \left\{ n_t = \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix} \right\}$$

for $t \in \mathbb{R}$ for $n = 2$ and $t \in \mathbb{C}$ for $n = 3$. This gives a decomposition via the spectral theorem that $G = KA^+K$ called the *Cartan decomposition*. This decomposition already tells us that the Haar measure is left- and right-invariant because this presentation of G is symmetric in this decomposition; K acts both on the left and the right. In fact, this is true in higher dimensions as generalized in the following theorem.

Theorem 2.3.2. *The isometry group of $\mathbb{H}^n$ is unimodular.*

We can compute what this Haar measure is for G. Because K does not change the radius and can act on the left and the right, we know that the Haar measure can be expressed simply as a function of the radius r for any element $ka_rk' \in G$.

First let us work for $n = 2$ so $G = \mathrm{PSL}(2, \mathbb{R})$ and $\mathbb{H}^2 = G/K$. Let $\mathfrak{a}$ be the Lie algebra of A, and $\mathfrak{s}$ be the Lie algebra of $K = \mathrm{SO}(2)$. A has natural coordinate r. We can take the exponential map in this decomposition and fix some r to get a map

$$\mathfrak{s} \oplus \mathfrak{a} \oplus \mathfrak{s} \to G, \qquad \theta \oplus \alpha \oplus \theta' \mapsto \exp(\theta) \cdot a_r \exp(\alpha) \cdot \exp(\theta')$$

which induces a linear map on the tangent space of G at a_r, which can be left-multiplied by a_r^{-1} to be $\mathfrak{g}$ the tangent space at the identity of G (its Lie algebra) to get a map

$$\theta \oplus \alpha \oplus \theta' \mapsto a_r^{-1} \exp(\theta) \cdot a_r \exp(\alpha) \cdot \exp(\theta').$$

We can now act by the adjoint action of a_r on θ which will show that these span $\mathfrak{g}$.

To finish computing the Haar measure, we can now choose a basis and compute the radial dependence of the Haar measure in that basis. We choose a standard basis

$$h = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \in \mathfrak{a}, \qquad \theta = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \in \mathfrak{s}, \qquad \sigma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Taking the adjoint action of a_r^{-1} on θ gives

$$a_r^{-1} \theta a_r = \begin{pmatrix} 0 & e^{-r} \\ -e^r & 0 \end{pmatrix}$$

and this is expressed in the basis as $x\sigma + y\theta$ where x and y solve the system of equations

$$x + y = e^{-r} \qquad \text{and} \qquad x - y = -e^{-r}$$

which is solved by $x = -\sinh(r)$ and $y = \cosh(r)$. Therefore, the map from $\mathfrak{s} \oplus \mathfrak{a} \oplus \mathfrak{s}$ to $T_{a_r}G \cong \mathfrak{g}$ is given by

$$a\theta \oplus bh \oplus c\theta \mapsto -a\sinh(r) \cdot \sigma \oplus bh \oplus (c + a\cos(r))\theta$$

or more generally, the map from $\mathfrak{s} \oplus \mathfrak{a} \oplus \mathfrak{s} \to \mathbb{R}\langle\sigma\rangle \oplus \mathfrak{a} \oplus \mathfrak{s}$ is given by

$$(a, b, c) \mapsto (-a\sinh(r), b, c + a\cosh(r).$$

This computes the Haar measure on G as $d(ka_rk') = |\sinh(r)|dkdrdk'$.

To extend the above to $\mathrm{PSL}(2,\mathbb{C})$ does not require much extra work. We must replace $\mathfrak{s}$ to be the Lie algebra of SU(2) which has more components than SO(2). This is decomposed as

$$\mathfrak{s} = \mathbb{R}\langle\theta\rangle \oplus \mathbb{R}\langle i\sigma\rangle \oplus \mathbb{R}\langle ih\rangle$$

which are the Pauli matrices (up to multiplication by i according to convention). What makes this so similar to the above computation is that a_r acts as above on the first summand and trivially on the third summand above. Therefore, we only need to compute its action on the middle summand, which is similar to the first. Let $\alpha = i\theta$ and $\beta = i\sigma$. As above, we must solve for coefficients x and y such that $x \cdot \alpha + y \cdot \beta = a_r^{-1}\beta a_r$ which gives the system of equations

$$x \cdot i + y \cdot i = e^{-r} \cdot i \qquad \text{and} \qquad x \cdot (-i) + y \cdot (-i) = e^{r} \cdot i$$

which is solved by $x = -\sinh(r)$. Therefore, this terms picks up an extra $\sinh(r)$ so we get that the Haar measure on G is $d(ka_rk') = |\sinh(r)|^2 dkdrdk'$.

2.3.2 Kleinian groups

We can now define a Kleinian group which is the key algebraic structure used to construct hyperbolic manifolds.

Definition 2.3.3 (Kleinian group)**.** A *Kleinian group* is a discrete, finitely generated subgroup of the isometries of $\mathbb{H}^n$. In dimension $n = 2$, this is a subgroup of $\mathrm{PSL}(2,\mathbb{R})$ and in dimension $n = 3$, a subgroup of $\mathrm{PSL}(2,\mathbb{C})$.

Kleinian groups correspond to the fundamental groups of hyperbolic manifolds, so they give algebraic invariants of the topology up to group isomorphism. What the algebraic version of Mostow rigidity will say is that any pair of Kleinian subgroups of the isometry group of $\mathbb{H}^n$, i.e. $\mathrm{PSL}(2,\mathbb{C})$ for $n = 3$, corresponding to a complete finite volume hyperbolic 3-manifold will not only be isomorphic, but conjugate. Therefore, we can take all the traces of the elements which will be left unchanged. These will form a field, a number field in fact, called the *trace field* of Γ and give another algebraic invariant of $M = \mathbb{H}^n/\Gamma$.

2.4 (G, X)-manifolds and the developing map

Given a hyperbolic manifold M, how can we exhibit M as $\mathbb{H}^n/\Gamma$? Futher on, in Section 2.7, we will prove the classification of simply connected space forms (Theorem 2.7.1), which will express a complete hyperbolic manifold as $\mathbb{H}^n/\Gamma$. First, in this section, we want to show an explicit way to do this by constructing the *developing map* which unfolds a geometric object in hyperbolic space. This concept will also extend to incomplete manifolds and the description of completeness will be characterized by whether or not when we unfold them, we achieve the entire hyperbolic space or not. That is, the failure of completeness will be measured by the geometric unraveling of the manifold failing to encompass all of $\mathbb{H}^n$.

This concept generalizes to classes of manifolds called (G, X)-manifolds which are locally modeled by some homogeneous space X with transition functions in a (Lie) group G.

Definition 2.4.1 ((G, X)-manifold)**.** Let X be a smooth, connected manifold and G be a subgroup of diffeomorphisms of X that act *analytically* on X. This means that for $g, g' \in G$, if there exists an open set $U \subset X$ in which g and g' agree, $g|_U = g'|_U$, then $g = g'$. A smooth manifold M is said to be a (G, X)-*manifold* if M has an atlas of open sets U_i that are modelled on X in the sense that there are diffeomorphisms $\varphi_i : U_i \to V_i \subset X$ to open subsets V_i such that the transition functions $\varphi_i \circ \varphi_j^{-1}$ are the restrictions of the action of some $g \in G$. Each (U_i, φ_i, V_i) is called a *chart*.

We say that two (G, X)-structures are equivalent if they are both contained in a further (G, X)-atlas called a *refinement*. Equivalently, if their union satisfies the (G, X)-structure requirements, it is such a refinement, so we only must verify this single amalgamation as a potential (G, X)-atlas.

Example 2.4.2 (Hyperbolic manifolds are (G, X)-manifolds). Any hyperbolic manifold is a (G, X)-manifold where X is $\mathbb{H}^n$ and G is the isometry group of hyperbolic space. This is to say we can have an atlas of open sets on a hyperbolic manifold each of which has a natural geometry induced by the hyperbolic metric preserved by the transition functions.

The important feature of the (G, X)-structure, which will allow us to defined the developing map, is the analytic condition. In particular, we will use this to unfold hyperbolic manifolds. Using the above notation, we denote $\gamma_{ij} = \varphi_i \circ \varphi_j^{-1}$ to be our transition functions, and on the (G, X)-atlas, these functions locally agree with elements of G. By the analytic condition, on some U_i, these must be constant, and therefore, γ_{ij} are locally constant. Notably, they may not be truly constant because $U_i \cap U_j$ may contain multiple disjoint components.

Consider some $x \in U_i \cap U_j$, and we can precompose φ_i with γ_{ij} so that we may assume that $\varphi_i(x) = \varphi_j(x)$ as points in X. Therefore, we can now glue together the functions φ_i and φ_j to a single function defined on $U_i \cup U_j$ using the analytic property. This would be the notion of unfolding M by the geometric pieces that are the U_i. However, we may have a problem of well-definedness if we continue to do this and eventually return back to our point x. An issue that could occur if we continue unfolding over open sets $U_1, \ldots, U_n$ such that $x \in U_1$ and $x \in U_n$. Suppose we trace a loop starting and ending at x, the values of the extended function from $M \supset \bigcup_{i=1}^{n} U_i \to X$ may not agree at x. The obstruction here would be non-trivial homotopy classes, so we realize that we must replace M with $\tilde{M}$, its universal cover.

We can now define the developing map as this analytic continuation, after fixing a basepoint, from $\tilde{M}$ to the model space X.

Definition 2.4.3 (Developing map)**.** For a (G, X)-manifold M and universal cover $\pi : \tilde{M} \to M$, let $x_0 \in U_0$ be a basepoint with $\varphi_0 : U_0 \to X$ given as a chart in the (G, X)-structure. We define the *developing map* as $D : \tilde{M} \to X$ that agrees with the analytic continuation of φ_0 along each path emanating from x_0 in a neighborhood of the path's endpoint. That is to say, $D = \varphi_0^y \circ \pi$ in an open neighborhood around $y \in \tilde{M}$, the end of the path.

Since $\pi : \tilde{M} \to M$ is a local diffeomorphism, $\tilde{M}$ carries an induced (G, X)-structure from M. Let σ be an element in $\pi_1(M)$ based at $x \in M$. We can lift σ uniquely to a map $\tilde{\sigma} : [0, 1] \to \tilde{M}$ such that $\pi(\sigma(0)) = \pi(\sigma(1)) = x$. Notably, for our charts φ_0 near x given by the (G, X)-structure, σ gives two charts around x by analytic continuation on $\tilde{M}$ and the projection map. This means there exists some g_σ such that $\varphi_0^\sigma = g_\sigma \varphi_0$ for φ^σ being the (G, X)-chart at x, given by going around the loop σ. This can be characterized using the *deck transformations* of the covering map. The holonomy equation is given by

$$D \circ T_\sigma = g_\sigma \circ D$$

for $T_\sigma : \tilde{M} \to \tilde{M}$ the deck transformation corresponding to $\sigma \in \pi_1(M)$ and D the developing map.[4] The set of such $g_\sigma \in G$ is a subgroup and is called the *holonomy*. It measures how the (G, X)-structure can change as one travels around non-trivial loops, which is the failure of the well-definedness analytic continuation without passing to the universal cover.

The holonomy can recover the (G, X)-structure of M exactly when M is complete. We further have a method to verify the completeness topologically. M is said to be a *complete* (G, X)*-manifold* if its developing map $D : \tilde{M} \to X$ is a covering map. Notably, when X is simply connected, that means it must be a homeomorphism. This will align with the discussion in Section 2.7, where given a complete metric of constant sectional curvature -1 on M, we will use the exponential map to give a covering map $\mathbb{H}^n \to M$, which according to the observations above, will be computed by the developing map.

2.5 Polyhedra

Many hyperbolic constructions hinge upon the definition of polyhedra, which are well-defined in any model space of constant curvature, (with a bit of care taken for the sphere).

[4]Deck transformations of a covering map $\pi : E \to B$ are maps (up to homotopy) that commute with π. For E the universal cover of B, this is naturally identified with $\pi_1(B)$, which is the fact being used here.

Here we develop the theory of hyperbolic polyhedra and demonstrate how to use them to build hyperbolic manifolds. More generalized handling of polyhedra in spherical, flat, and hyperbolic contexts can be found in Bonahon [Bon09]. The goal will be to utilize that hyperbolic manifolds are expressible as $\mathbb{H}^n/\Gamma$, a quotient of hyperbolic space by a Kleinian group (a subgroup of the isometry group that acts discontinuously). This means M has a fundamental domain that will be a polyhedron, and some identifications of the sides to study it similarly to how a torus is viewed as $\mathbb{R}^n/\mathbb{Z}^n$ with fundamental domain $[0,1]^n$ with opposite sides identified. We will construct this in Subsection 2.6.1. This example motivates the idea that compact hyperbolic manifolds should have a bounded fundamental domain, while non-compact manifolds will have a fundamental domain that extends to infinity, or an ideal vertex (see Definition 2.5.8). More complex hyperbolic manifolds will be realized as the gluing of multiple polyhedra, not necessarily bounded, along isometries of their sides. If a side is identified to itself, it is a boundary.

First, a definition of convexity is needed to be extended for all the model spaces.

Definition 2.5.1 (Convexity)**.** Let X denote one of the complete space forms $\mathbb{R}^n, \mathbb{H}^n$ or S^n as detailed in Theorem 2.7.1. Points $p, q \in X$ are called proper if there exists a unique geodesic adjoining them. For $\mathbb{R}^n$ or $\mathbb{H}^n$, these can be arbitrary, but for S^n, this means they cannot be antipodal points. A connected region $\Omega \subset X$ is convex if for every pair of proper points $p, q \in \Omega$, all the points along the geodesic connecting them are also contained in Ω. This aligns with the natural definition for Euclidean space.

Remark 2.5.2. This definition may differ from the literature in that we required that convex sets are connected. If Ω were not assumed to be connected, as is in Bonahon [Bon09], then a pair of antipodal points $\{p, -p\} \subset S^n$ would be considered convex. All convex sets will be assumed to be connected, as this is the only counter-example to be avoided.

Definition 2.5.3 (Planes)**.** A plane of dimension k in the ambient space X being $\mathbb{R}^n, \mathbb{H}^n$ or S^n is defined as

- $X = \mathbb{R}^n$: an arbitrary plane of dimension k in the Euclidean sense.
- $X = \mathbb{H}^n$: a k-dimensional half-plane perpendicular to the hyperplane $x^n = 0$ or a k-dimensional half-sphere with center on the plane $\{x^n = 0\}$.
- $X = S^n$: a great sphere of dimension k, that is, a model is given by

$$\{(x^1, \dots, x^{k+1}, 0, \dots, 0) \in S^n \subset \mathbb{R}^{n+1} : \sum (x^i)^2 = 1\}$$

  and all *rotations* of it as acted on by the group $\mathrm{O}(n+1)$. Alternatively, $S^n \cap P$ for P a $(k+1)$-dimensional plane through the origin.

Definition 2.5.4 (Dimension)**.** The dimension of a convex set $\Omega \subset X$ will be the minimal dimension k of a k-plane that it lies inside.

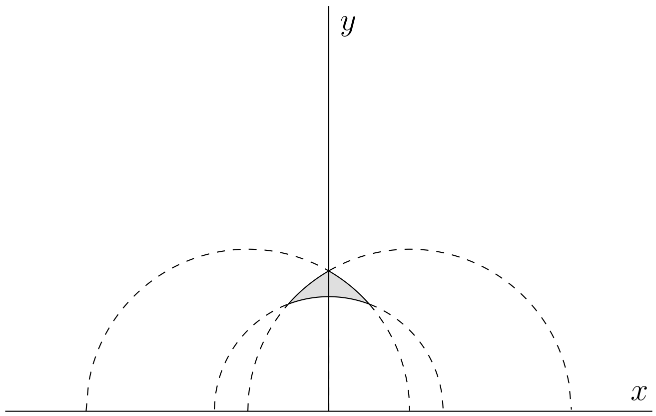

Figure 2.2: An example of a bounded triangle in $\mathbb{H}$.

Definition 2.5.5 (Sides)**.** A side of a convex set Ω is a maximally convex subset of the topological boundary $\partial\Omega$.[5]

Definition 2.5.6 (Polyhedron)**.** A polyhedron is a closed convex region $\Omega \subset X$ such that its boundary $\partial\Omega$, considered as a collection of its sides, is locally finite in X.

Example 2.5.7 (Non-example)*.* The locally-finite condition is a concise way to rule out *curved* sides. For example, the unit disk $D^2 \subset \mathbb{R}^2$ is not a polyhedron because its sides would be every point of $S^1 = \partial D^2$ bounding it. Since every point on the boundary is an extreme point, each point is a different side. Therefore, the set of sides is not locally finite. This same argument works for any curved convex region.

Definition 2.5.8 (Ideal Vertex)**.** Suppose a polyhedron has two geodesics that are maximally extended on one side and meet at a point on the sphere at infinity. This point is said to be an ideal vertex. This is a slight abuse of notation, as this point does not lie on the hyperbolic plane, so it is not a part of the polyhedron, but it is useful to think about these as extended to the boundary nonetheless.

Example 2.5.9*.* Polyhedra are notably allowed to be infinite as the closed condition just means they contain their boundary. The closed half plane $\{x : x^n \geq 0\} \subset \mathbb{R}^n$ is a polyhedron in $\mathbb{R}^n$ as is the region $\{z : 0 \leq \mathrm{Re}(z) \leq 1, |z| \geq 1\} \subset \mathbb{H}$, an infinite hyperbolic polyhedron (see Figure 2.3). By the closed property, infinite polyhedra are impossible in the round setting as S^n is compact. Furthermore, polyhedra in S^n have finitely many sides since locally finite in a compact space is finite.

Definition 2.5.10 (Ideal Polyhedra)**.** For $\mathbb{H}^n$, there is a special class of polyhedra known as ideal polyhedra. These are non-compact and have sides which are maximal geodesics

[5]Note that if a convex set were to be allowed to be not connected as in the definition from Bonahon [Bon09], then any geodesic connecting two antipodal points would have the antipodal points together as a single side and be a one-sided convex set with a disconnected boundary, which is absurd.

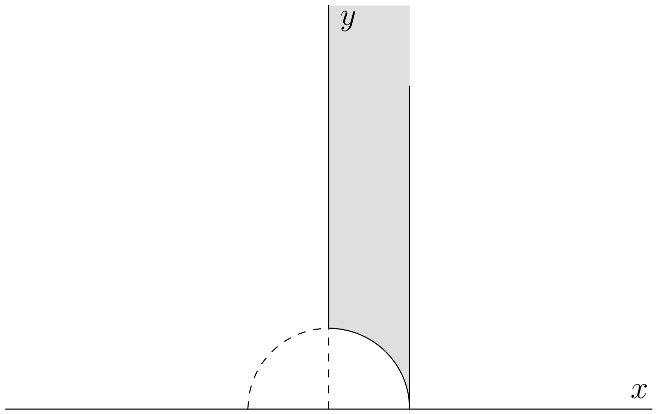

Figure 2.3: The infinite polyhedron described in Example 2.5.9 shown here is above the circular arc, with sides the circular arc and vertical lines. It has two ideal vertices at the points 1 and at ∞, while the point i is not an ideal point, as neither of the geodesics connected to it are maximal as shown by the dashed extensions.

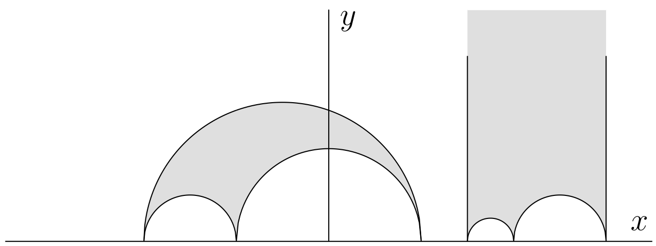

Figure 2.4: Ideal polyhedra in $\mathbb{H}$. The region between the three semi-circular geodesics on the left is an ideal triangle. The region above the two semicircles and between the vertical lines on the right is an ideal quadrilateral. The area of the left ideal triangle is π and the right ideal quadrilateral is 2π.

whose boundary in $\overline{\mathbb{B}^n}$, $\overline{\mathbb{H}^n}$, or any model closed to include the sphere at infinity, intersected with the sphere at infinity is finite. Informally, they have *only vertices* which lie on the sphere at infinity. The finiteness of the intersection with the sphere at infinity rules out infinite volume polyhedra – we do not want polyhedra to have a "side" at infinity. See Figure 2.4 for examples of ideal polyhedra and Figure 2.5 for a non-example.

Definition 2.5.11 (Ridges)**.** Consider P a polyhedron of dimension n. Every side of P is itself a polyhedron of dimension $n-1$. Let $\mathcal{S}(P)$ be the set of sides of P, and similarly for S a set of polyhedra, let $\mathcal{S}(S)$ be the set of sides of all $P \in S$. Define a ridge to be some polyhedron of dimension at most $n-2$ which is in $\mathcal{S}(\cdots\mathcal{S}(P))$. Denote $\mathcal{S}_k(P)$ to be $\underbrace{\mathcal{S}(\cdots\mathcal{S}}_{n-k \text{ times}}(P))$ the ridges of dimension k.

Example 2.5.12. Consider a regular pentagon with five right angles in the hyperbolic plane. Gluing four of these together yields the pair of pants manifold, a building block of compact Riemann surfaces. The gluing is shown below in Figure 2.6 with unmarked sides forming the boundary. The right angles ensure that this has a well-defined manifold structure, as opposed to an orbifold if the angles did not add up.

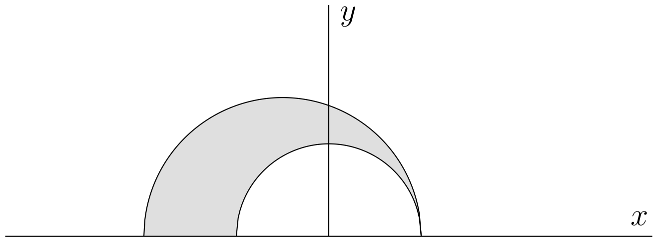

Figure 2.5: The shaded region is not an ideal polyhedron in $\mathbb{H}$, as it has a "side" at infinity. Its area is infinite.

If instead of regular pentagons, we prescribe right-angled pentagons with matching lengths of the gluing sides, we can create pairs of pants with any triple cuff lengths. The lengths of the circles forming the boundary are hyperbolic invariants, notably, since there is no rigidity in dimension two. This gives a moduli space of pairs of pants given by $\mathbb{R}^3_{>0}$.

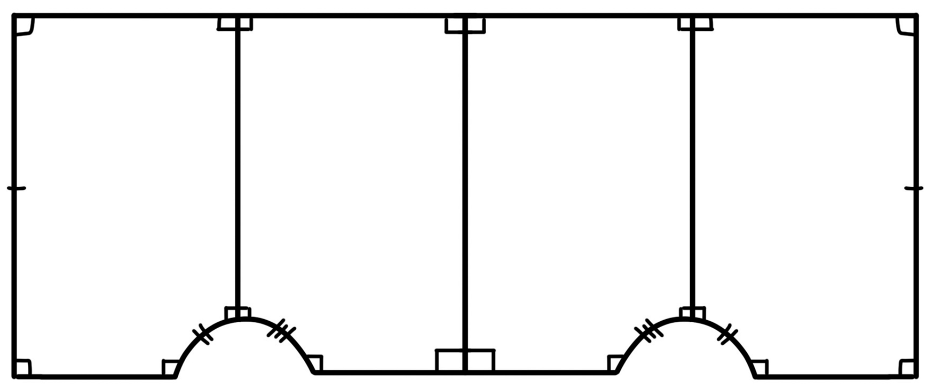

Figure 2.6: Gluing four right-angled pentagons to construct a pair of pants. The sides not glued form the boundary. The lengths of the three circles forming the boundary are hyperbolic invariants. The outer horizontal cut would be akin to cutting open your pants from your left ankle to left hip. The middle cuts would be the interior of each leg from one inner ankle to the other. Figure created by Kalia Firester.

Example 2.5.13 (Riemann surfaces). A fundamental construction in the theory of hyperbolic Riemann surfaces is colloquially known as the *pair of pants decomposition*, and its roots lie in the classification of surfaces which is due to building up a Riemann surface via a Morse function. We will not detail this construction and we refer the reader to Donaldson [Don11], Chapter 2.

A compact Riemann surface of genus $g \geq 2$ can be constructed with $2g - 2$ hyperbolic pairs of pants as constructed above in Example 2.5.12. This can be seen by gluing two pairs of pants together to get a genus 2 surface, (see Figure 2.7), and every additional genus requires two more pairs of pants. In order to glue them together according to their boundary cuffs, we need the cuff-lengths to come in pairs.

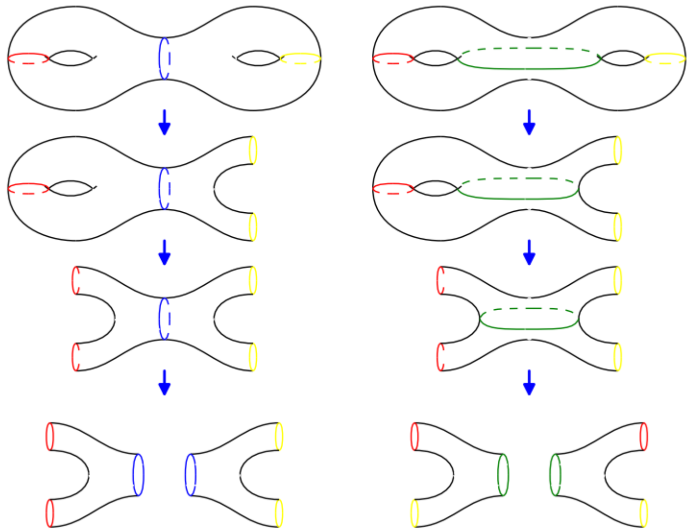

Figure 2.7: A genus 2 Riemann surface constructed by gluing hyperbolic pairs of pants. Image created by Jean Raimbault and accessed from https://en.wikipedia.org/wiki/Pair_of_pants_(mathematics).

As described in the above example, these cuff-lengths can be prescribed as any triple of positive real numbers. Therefore, we have $3g-3$ dimensions to choose these pairs of pants. Furthermore, the cuffs can be glued by rotating the cuffs, so there is an angle component associated to each gluing that is parameterized by an arbitrary real number. There are $3g-3$ pairs of cuffs to be glued, and there is a further $3g-3$ dimensional space of angles to glue along.

The above construction shows a $6g-6$ real-dimensional parameter space of hyperbolic structures on Riemann surfaces of genus g. We will denote this space $\mathcal{T}_g$, and we have shown it is contractible.[6] There is an action of the mapping class group, which is the group of diffeomorphisms modulo smooth isotopy, on $\mathcal{T}_g$, and the quotient is the moduli space $\mathcal{M}_g$ which will have dimension $3g-3$. The mapping class group captures discrete diffeomorphism classes. For further details on this topic, we refer the reader to Hubbard [Hub06] as a robust exposition of the field.

There are two other main constructions of $\mathcal{M}_g$, one using geometric invariant theory, and another by Hodge theory. This was rigorously done by Deligne and Mumford [DM69], and they further developed a *compactification* (algebraically proper) $\overline{\mathcal{M}_g}$ in the algebraic setting. The notion that $\mathcal{M}_g$ is not a manifold, but an orbifold, is algebraically expressed that $\mathcal{M}_g$ is not a scheme, but a stack.

The dimension $3g-3$ of $\mathcal{M}_g$ was known for quite some time, with many foundational results in algebraic geometry assuming the existence of a moduli space before it was rigorously constructed. The dimension can be computed quite easily using the Hilbert scheme, for example. For further reading, we suggest Harris and Morrison [HM98].

[6] This space is known as *Teichmüller space*, but Teichmüller was a huge Nazi, so perhaps we ought to avoid this name to only commemorate the math, and not the man.

Example 2.5.14 (Seifert-Weber dodecahedral space). We can use a method of continuity to find polyhedra that glue well. Consider a hyperbolic dodecahedron. This can be found by taking, in the Poincaré disk model $\mathbb{B}^3$, points all of Euclidean distance r from the origin that are the vertices of a Euclidean dodecahedron and taking their hyperbolic convex hull. As r varies from 0 to 1, there will be a unique radius where we will be able to glue this hyperbolic dodecahedral polygon along its boundary to get a smooth hyperbolic manifold. We can show that when $r = 1$, the ideal dodecahedron will have dihedral angles that are $\pi/3$, and when $r = \varepsilon$, its dihedral angles will be quite close to the Euclidean dihedral angle of $2\,\mathrm{atan}(\gamma) \approx 116.5^\circ$ for $\gamma = \frac{1+\sqrt{5}}{2}$, the golden ratio. Notably, the gluing condition of five dihedral angles adding to 2π will be at 72° degrees, which falls between these values. Therefore, there is indeed a unique good value of r that will allow such a gluing to satisfy the manifold criteria. The gluing here is given by identifying opposite sides by a $3\pi/5$ twist. This space is called the Seifer-Weber dodecahedral space, and is in fact not a Haken manifold, meaning it does not contain an orientable, incompressible surface.

Haken manifolds are historically interesting because Thurston first proved a hyperbolization theorem for Haken manifolds, later generalized in the geometrization theorem. We point the reader to McMullen [McM92] for more details on the geometrization of Haken manifolds.

Theorem 2.5.15 (Thurston's hyperbolization theorem). *If M is a compact, irreducible, atoroidal, Haken manifold with boundary satisfying $\chi(\partial M) = 0$, then $M^\circ = M \setminus \partial M$ has a complete hyperbolic structure with finite volume.*

Atoroidal here means that M does not contain an embedded, non-boundary parallel, incompressible torus.

Example 2.5.16 (Tetrahedron outside the sphere at infinity). We can use the above example to motivate other gluings of regular polyhedra. We know that if we tried to glue a tetrahedron to itself by identifying the boundary sides, it would never yield a manifold structure, as the Euler characteristic does not vanish (see Proposition 2.6.2). Furthermore, even with ideal angles, dihedral angles will not add up. A key insight is to not stop at a tetrahedron with its vertices on the boundary sphere at infinity, but to take one with its vertices outside.

The problem is that now the part of the tetrahedron that lies inside $\mathbb{B}^3$ intersects the sphere at infinity in a large region, so it is very far away from being finite volume. We can take the boundary of the intersection with the sphere at infinity and cap it off geodesically with hyperbolic planes meeting at right angles. This will yield an ideal polyhedron inside $\mathbb{B}^3$ that is of finite volume. We can size the tetrahedron to have dihedral angles quite small of $\frac{\pi}{n}$ and $\frac{\pi}{2}$ between the hyperbolic planar caps using continuity. This is visualized in Figure 2.8.

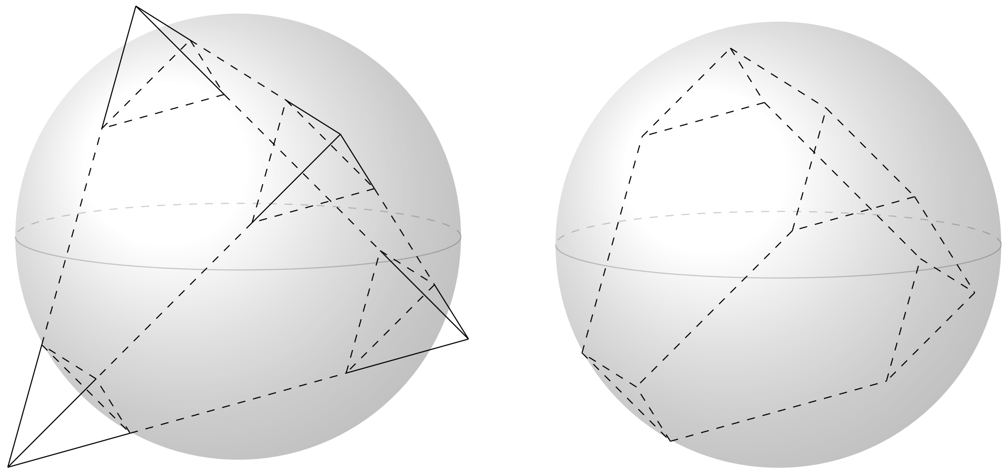

Figure 2.8: In the Klein model $\mathcal{K}^3$, we have a tetrahedron outside the unit disk and to construct a hyperbolic manifold, we take its intersection with $\mathcal{K}^3$ and cap off the components, so that the resulting space is a polyhedron with finite volume. We size the tetrahedron so the dihedral angles between the tetrahedron sides are π/n, and the diheral angles between the caps are $\pi/2$. The right image shows the ideal hyperbolic polyhedron inside $\mathcal{K}^3$. The dashed lines lie inside the sphere $\mathcal{K}^3$, and the bold lines lie outside. The edges of the fundamental domain come in two types, denoting the two types of dihedral angles. The first are the longer lines that are edges of the original tetrahedron. The second are the shorter ones forming the caps around each vertex.

We can now associate a group Γ generated by all reflections through the sides of the polyhedron. To finish, we take a finite index subgroup that is torsion free $\Gamma' \subset \Gamma$ to get $\mathbb{B}^n/\Gamma'$, a finite volume hyperbolic 3-manifold.

Example 2.5.17 (Orbifolds). Reckless gluing of polyhedra can have bad orbifold singularities. See Proposition 2.6.1 for conditions to glue polyhedra and get a manifold. We could instead consider this a feature, not a bug, as these are interesting spaces in their own right and do come up naturally, such as in moduli spaces. As mentioned above, for Riemann surfaces, the modueli space $\mathcal{M}_g$ is an orbifold, not a manifold. This can be rephrased as giving only the existence of a *coarse*, but not *fine*, moduli space of Riemann surfaces.

Expanding to hyperbolic orbifolds, we can construct many examples by taking polyhedra and gluing pairs of their faces together along isometries. Any polyhedron can be turned into a closed orbifold in this manner by taking two copies of it and identifying the faces along the identity maps. If we take an ideal tetrahedron and perform this procedure, we get a hyperbolic orbifold that is not a manifold because it does not satisfy Poincaré duality.

Example 2.5.18 (Bad gluing of ideal triangles). If we take an ideal triangle and glue it to itself along the canonical identifications of the boundary, the resulting manifold is a triply-punctured sphere, and is geodesically complete. However, this property generally fails if the gluing is not along the canonical identifications. We work in the upper half-plane model, and will shift only a single edge of the triangle. In Figure 2.9 below, we have shown two ideal triangles already glued at the vertical edge. We glue the two maximal geodesics that are the semi-circles via the canonical gluing, which is reflection. However, we glue the left most vertical line to the right most vertical line by vertical multiplication by two. We

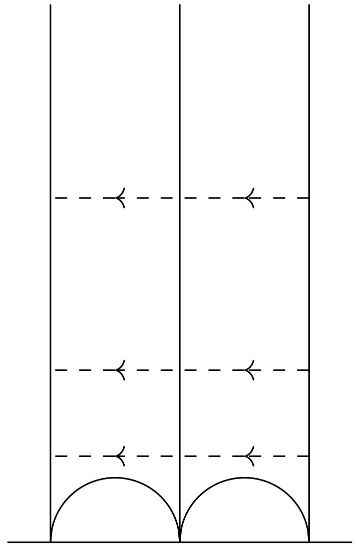

Figure 2.9: A path of finite length escaping to infinity.

know that the distance between two vertical points $x + iy$ and $x + iy'$ is $\log(y/y')$ from integrating the metric $\frac{1}{y}$. Therefore, multiplication by a scalar is an isometry. We can now take the curve that is always moving horizontally and to the left. Each time it hits the left-most vertical line, it jumps to twice its height and comes back on the right-most vertical line. The length as it travels to the left is $\frac{2}{y^2}$ assuming that the outer most vertical lines are at real parts ± 1. The distance is computed by integrating $\frac{1}{y}$ as x goes from 1 to -1. Therefore, we get the length of the curve starting at height y to be

$$\frac{2}{y^2} + \frac{2}{(2y)^2} + \frac{2}{(4y)^2} + \cdots = \frac{2}{y^2} \sum \frac{1}{2^{2n}} < \infty$$

is finite, so we have found a curve that escapes to infinity in finite time, showing this space is incomplete. In fact, this curve could be shortened by replacing these segments with the geodesics connecting the endpoints, but this would have diminishing effect as $y \to \infty$ and the geodesics become better approximated by horizontal lines. We can also see the incompleteness through the developing map. In this case, as we unfold each ideal triangle, the vertical height doubles as we move to the left. This has a geometric picture; instead of having the dashed path in Figure 2.9 jump back, we could continue unfolding each triangle and resize it appropriately. This is shown in Figure 2.10.

2.6 Constructing hyperbolic 3-manifolds from polyhedra

Consider a set $P_1, \ldots, P_m$ of polyhedra in $\mathbb{H}^n$, and let $P = \bigsqcup P_i$. Define a gluing map that identifies pairs of sides, potentially gluing a side to itself via the identity, giving rise to $q : P \to M$ the quotient map. If M is a well-defined manifold, this gives a hyperbolic structure on the interior of M since q is a homeomorphism on the interior, so the metric can be

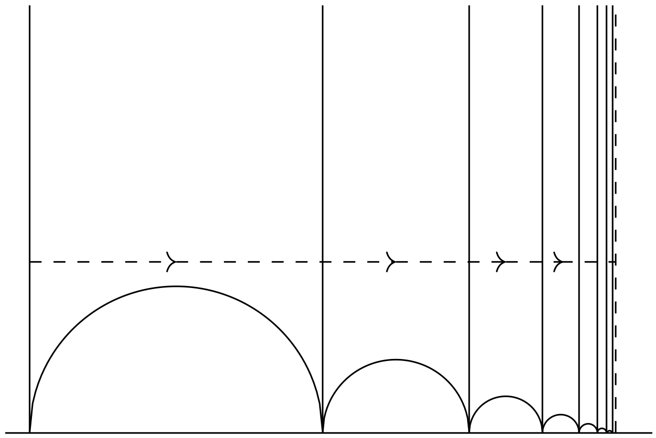

Figure 2.10: Each ideal triangle is glued to the next, unfolding the "bad gluing". Since each ideal triangle has half the width of its left neighbor, this unfolding never attains the right most dashed line, and has infinitely many ideal triangles squeezed up until it. The dashed path shows the incompleteness, as it has a finite length, but approaches *infinity* represented by the vertical dashed line in finite time.

carried over from P to M on this region. Using this description, if M is a well-defined hyperbolic manifold, it can be understood by a fundamental domain in $\mathbb{H}^n$ by taking some representation of P_1 embedded, and if a side is identified to P_i, draw that polyhedron in $\mathbb{H}^n$ sharing this side and continue for all P_i.

Denote M° as $q(P \setminus \partial P)$ the image of the interior, which does have a well defined hyperbolic structure. The question of when this extends to be a global structure on M can be reduced to a local one as follows. There are many other gluing criteria and properties that can help understand hyperbolic manifolds and orbifolds. For further constructions on hyperbolic polyhedra and gluing, we refer the reader to Lackenby [Lac00].

Proposition 2.6.1. *Suppose for every $x \in M$ there exists a homeomorphism $\phi : U_x \to B_\varepsilon(0) \subset \mathbb{B}^n$ mapping x to 0 that is an isometry on each component of $U_x \cap M^\circ$. Then M has a hyperbolic structure.*

Proof. For $x \in M^\circ$, this is always true as ε can be taken to be smaller than the distance to any of the boundary sides of the P_i containing x, and then U_x is given by a hyperbolic chart centered at x. Consider that for $x \in M \setminus M^\circ$, by shrinking ε, the restriction of U_x to each component of M° can contain x. That is, x lies on some ridge and consider ε small enough so that U_x only intersects the sides that are adjacent to x. This will give the necessary hyperbolic chart to define a hyperbolic structure globally. Consider an arbitrary preimage $q(\tilde{x}) = x$. A small ball around $\tilde{x}$ can be mapped through ϕ_x to $\mathbb{B}^n$ by ϕ_x. That is, for each $\tilde{x} \in P$, the map ϕ_x gives

$$h_{\tilde{x}} : B_\varepsilon(\tilde{x}) \to B_\varepsilon(0) \subset \mathbb{B}^n, \quad h_{\tilde{x}}|_{M^\circ} = \phi_x \circ q. \tag{2.17}$$

The only verification left to ensure these charts ϕ_x define hyperbolic charts covering M is that the transition functions are isometries, that is, $\phi_y \circ \phi_x^{-1}$ is an isometry of regions in $\mathbb{B}^n$.

This is immediately true for all $x \in M^\circ$ by considering $x \in P_i$ an interior point and the restriction of the hyperbolic metric on P_i to the charts around close points x and y. Let X be a component of $U_x \cap U_y$. Take a path between any two $x, y \in X \cap M^\circ$ which avoids the ridges (see Definition 2.5.11). For any $z \in \phi_x(q(\partial P) \cap X) \subset B_\varepsilon(0)$ that is not the image of a point lying on a ridge, the map $\phi_y \circ \phi_x^{-1}$ must be an isometry on some neighborhood of z. Take two distinct points $z_1, z_2 \in P$ that are preimages of $\phi_x^{-1}(z)$ under q. If these do not exist, then it means the side that contains z was identified to itself, and it is a boundary point of M, so it can be ignored. This means that $z_1 \in F_1$ and $z_2 \in F_2$ for F_1, F_2 some sides of P_1 and P_2 that are identified isometrically via the map $k : F_1 \to F_2$. Similarly, $x_1, y_1 \in F_1$ are preimages of x and y in the first face, and $x_2, y_2 \in F_2$ the corresponding images in F_2.

The map k can be extended to an isometry of all $\mathbb{H}^n$ in a well-defined manner by enforcing that k composes with the h_{x_i} maps, that is $h_{x_2} \circ k = h_{x_1}$. That is, $h_{\tilde{x}}$ can be extended to all of $\mathbb{H}^n$ by considering an isometry that takes $\tilde{x} \in P_i$ to the point $0 \in \mathbb{B}^n$ and extending via the identity since the map $h_{\tilde{x}}$ will then just be the restriction to the small ball $B_\varepsilon(0) \subset \mathbb{B}^n$. Therefore, we have the commutative diagram given by these charts

$$\begin{tikzcd} \mathbb{B}^n \arrow[r, "h_{x_1}^{-1}"] \arrow[d, "h_{x_2}^{-1}"'] & B_\varepsilon(x_1) \subset F_1 \arrow[d, "h_{y_1}"] \arrow[dl, "k"'] \\ B_\varepsilon(x_2) \subset F_2 \arrow[r, "h_{y_2}"'] & \mathbb{B}^n \end{tikzcd} \quad \rightsquigarrow \quad \begin{tikzcd} \mathbb{B}^n \arrow[r, "h_{x_1}^{-1}"] \arrow[d, "h_{x_2}^{-1}"'] & \mathbb{H}^n \arrow[d, "h_{y_1}"] \arrow[dl, "k"'] \\ \mathbb{H}^n \arrow[r, "h_{y_2}"'] & \mathbb{B}^n \end{tikzcd} \tag{2.18}$$

which show that around z, the composition of $\phi_y \circ \phi_x^{-1}$ is an isometry. These therefore give M local charts that are isometric to open sets of $\mathbb{B}^n$ with isometric transition functions, so M is a hyperbolic manifold. □

The above argument shows that we get a manifold structure from a polyhedra gluing when the boundary of the polyhedra glue together to avoid orbifold singularities where the local structure is $\mathbb{R}^n$ and not $\mathbb{R}^n/C_d$ for C_d the cyclic group of order d. This is generalized to other CSC geometries by the Poincaré polyhedra theorem 2.6.4 below. This generalized theorem gives conditions that in dimension three simplify greatly. In particular, in dimension three, it is very easy to verify if we get a manifold, the only check that we must make is that the resulting space has vanishing Euler characteristic. This can be phrased simply by verifying that M satisfies Poincaré duality in the following Proposition 2.6.2.

Proposition 2.6.2. *Let $P_1, \ldots, P_v$ be three-dimensional polyhedra such that the number of faces with K sides is even. Pick an identification of the sides: a matching of orientation-reversing isometries between the faces. A priori, this gives a three-dimensional CW-complex. This is a manifold if and only if it has Euler characteristic 0.*

Proof. If the structure is a manifold, by Poincaré duality, the rank of the degree one and two cohomology groups are equal and therefore the Euler characteristic vanishes.

For the other direction, we must show that a vanishing Euler characteristic gives a manifold structure. We first give a bound on the Euler characteristic based on the links of all vertices. We use some constructions on a simplicial complex. In a simplicial complex Σ, we define the link of a simplex $\sigma \subset \Sigma$ notating it as $\mathrm{lk}(\sigma)$. Let $\tau_1, \ldots, \tau_n$ be the simplices containing σ. We can define σ_i to be the simplex opposite σ in τ_u. The link $\mathrm{lk}(\sigma)$ is given as the simplicial subcomplex by all the σ_i.

Consider a filled in triangle and σ be a vertex. There are τ_1, τ_2 the edges with one side σ and the other two vertices are σ_1, σ_2. There is also the 2-cell τ_3 that is the triangle and then σ_3 is the edge across from the specified vertex.

Let's assume by subdivision that all the polyhedra are tetrahedra. We can compute the Euler characteristic as the alternating sum of the different dimension cells, so let's label these v, e, f, t for the number of vertices, edges, faces, and tetrahedra. The Euler characteristic is $v - e + f - t$.

For each vertex v_i, consider $\mathrm{lk}(v_i)$, the link of v_i. Because we subdivided to all tetrahedra, we have relations among the number of k-cells. Each tetrahedron has 4 faces, but X has only half this amount since they are glued in pairs, so we know that $f = 2t$.

We consider the link of each vertex. If we take all these simplicial subcomplexes combined, we realize that every edge accounts for two vertices in these links of vertices. Similarly, each face has three edges surrounding it and each tetrahedron has four faces. This allows us to compute the sum of the Euler characteristics of the links (which are surfaces)

$$\sum_{i=1}^{v} \chi(\mathrm{lk}(v_i)) = 2e - 3f + 4t.$$

We now compute the following formula for $\chi(X)$ as

$$\begin{aligned}
\chi(X) &= v - e + f - t \\
&= v - \frac{1}{2}(2e - 2f + 2t) \\
&= v - \frac{1}{2}(2e - 3f + 4t) && \text{using } f = 2t \\
&= v - \frac{1}{2}\sum_{i=1}^{v} \chi(\mathrm{lk}(v_i)) \geq 0.
\end{aligned}$$

The last inequality uses that the Euler characteristic of S^2 is two, easily computed using that S^2 can be made from a single vertex and a single 2-cell. Any other surface has Euler characteristic strictly less than two. In general, $\chi(\Sigma_g) = 2 - 2g$, for g the genus.

What is left to be shown is the following claim, that X is a manifold if and only if every link of a vertex is a sphere S^2, (homeomorphically, but even homologically suffices as we only care that its Euler characteristic is 0). This will cause the above computation to be sharp, so $\chi(X)$ will vanish. Any gluing pattern in two dimensions always forms a manifold. From this, we use that the link of a vertex has an induced gluing structure from X which is a compact manifold, namely a genus g surface. We must have that X is locally simply connected, Σ_g must be simply connected which implies $g = 0$. By assumption, the above can only vanish if each link is a sphere. The following lemma therefore completes the result.

Lemma 2.6.3 (Spherical links)**.** *Let X be a simplicial complex defined by a polyhedra gluing. If the link of every simplex of dimension p is homeomorphic to S^{n-p-1}, namely a sphere of the proper dimension, then X is a manifold.*

Proof. Every point $x \in X$ has a neighborhood given by the product of a disk of dimension p with a cone over a link of a simplex. Let x lie in a p-simplex, and we take the link of said simplex, which is the *cross-section* of a neighborhood of σ. Furthermore, the simplex σ is homeomorphic to a disk D^p. The cone over the link, which by assumption is the sphere S^{n-p-1}, is D^{n-p}. Therefore, a neighborhood of x is $D^p \times D^{n-p}$, which is homeomorphic to D^n, giving a local chart. □

Geometrically, the obstruction given above for a link being Σ_g for $g \geq 1$ is that X must be locally simply connected to be a manifold. The above lemma showed that the neighborhood of such a vertex will be homotopic to the link, so any open set around it has π_1 the same as the link. Therefore, the presumption of locally simply connected forces each link to be S^2, and therefore the Euler characteristic vanishing enforces that X is locally simply connected. □

To extend this to higher dimensions, there are more verifications necessary on the higher codimension ridges. The result is summarized by the Poincaré polyhedra theorem below.

Theorem 2.6.4 (Poincaré polyhedra theorem)**.** *Let M be the space obtained by gluing finitely many totally geodesic compact polyhedra (spherical, Euclidean, or hyperbolic) P_i along pairwise isometric identifications of their sides. If for each codimension two ridge ϕ, the dihedral angles add up to 2π and the composition of the gluing isometries around ϕ are the identity on ϕ, then M has a (spherical, Euclidean, or hyperbolic) orbifold structure. Furthermore, if the links in codimension three and higher are simply connected, then M is a manifold.*

Example 2.6.5 (Limit of pair of pants). Going back to one more example of the geometric topology. A pair of pants is uniquely determined by the lengths of the three cuffs. As these shrink to zero, it becomes three cusps. This has the structure of the triply-punctured sphere with a unique hyperbolic metric. As a Riemann surface with negative Euler characteristic, it must be hyperbolic, and since the automorphisms of a sphere are triply transitive, it is unique. This is achieved by gluing two ideal triangles along the canonical identifications of their boundaries.

Example 2.6.6 (Pants whose cuffs do not shrink). Topologically, the pair of pants without the boundary is the triply-punctured sphere, but the hyperbolic structures do not match. This is seen because the pair of pants has geodesics of length bounded below on each cuff. Like the triply-punctured sphere, topologically, the interior of the pair of pants can be constructed by gluing two ideal triangles, but not along their canonical boundary identifications. Such a “bad gluing” was shown in Example 2.5.18. The picture is that each vertex

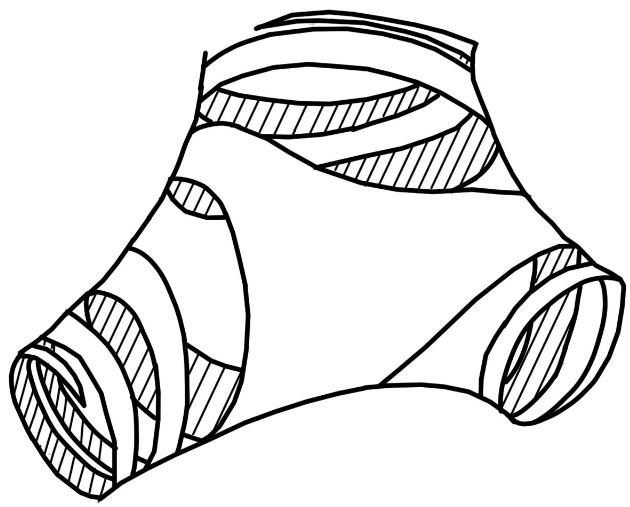

Figure 2.11: The interior of a pair of pants can be expressed as two ideal triangles glued along non-canonical identifications of their boundary sides, depending on the lengths of the cuffs. Figure created by Kalia Firester.

of the ideal triangles no longer goes to infinity, but wraps infinitely many times around each cuff, with closer and closer winding. This is shows below in Figure 2.11.

2.6.1 Dirichlet domains

Our goal in this subsection is to complete the picture of how to translate among the perspectives on hyperbolic manifolds. Given M presented as $\mathbb{H}^n/\Gamma$, we want to express M as a polyhedral gluing. We can take some $x \in M$ as a basepoint and lift it to an arbitrary $x_0 \in \mathbb{H}^n$. The key construction is to consider the orbit $\Gamma\, x_0 \subset \mathbb{H}^n$ which is a discrete collection of points indexed by $\gamma \in \Gamma$. The polyhedra gluing will be defined by taking the hyperbolic *Voronoi region* around x_0. This is the collection of points $y \in \mathbb{H}^n$ such that $d_{\mathbb{H}^n}(x_0, y) \leq d_{\mathbb{H}^n}(\gamma(x_0), y)$ for any $\gamma \in \Gamma$. What we imagine here is that each $\gamma(x_0)$ is the center of a region of points that are closest only to it. The gluing here will be given via identification of the sides of this domain via the isometries $\gamma \in \Gamma$, which is exactly the projection map $\pi : \mathbb{H}^n \to M$.

Definition 2.6.7 (Dirichlet domain)**.** For $x_0 \in \mathbb{H}^n$ and Γ a Kleinian group, the Dirichlet domain D is defined by taking the orbit of x_0 in $\mathbb{H}^n$ and letting D be the hyperbolic Voronoi region around x_0:

$$D = \{x \in \mathbb{H}^n : d_{\mathbb{H}^n}(x, x_0) \leq d_{\mathbb{H}^n}(\gamma(x), x_0)\, \forall\, \gamma \in \Gamma\}$$

which are the points closer to x_0 than to any other point in its orbit.

Theorem 2.6.8 (Dirichlet domain is a fundamental domain)**.** *The Dirichlet domain D is a connected, convex, locally finite fundamental domain for the complete, compact hyperbolic manifolds $M = \mathbb{H}^n/\Gamma$ with geodesic boundary.*

More generally, if Γ is an arbitrary discrete group, the theorem holds if $\mathrm{Stab}_\Gamma(x_0) = \mathrm{Id}$.

Proof. There are a couple of verifications necessary to show that this is indeed a polyehdral gluing schema. Firstly, this region is convex by the triangle inequality. We need to show that there are only finitely many sides, which is really a two-part problem: first we must show that only finitely many points $\gamma(x_0)$ that share points in $\mathbb{H}^n$ of minimal distance to both of them, and second, we must verify that the boundary consists of geodesic sides. The second point is quite immediate as the locus of points equidistant from any two points is a totally geodesic hyperplane. Therefore, the Voronoi region around x_0 is the intersection of half-spaces on each side of the geodesic boundary between x_0 and its neighbors, so it is a polyhedron if and only if there are only finitely many neighbors. Let us name this region A_0.

By construction, the interior of A_0 is the complement of a measure 0 set on M, so it has the same finite volume as M. This eliminates the case of the Voronoi region around x_0 containing an open region of the boundary at infinity, so it will be a closed polyhedron with potentially finitely many ideal vertices.

To verify locally finite, we must show that for any compact K contained in A_0, we have that $\gamma(K) \cap A_0 \neq \emptyset$ for only finitely many γ. It is further sufficient to check this on a disk centered at x_0 of radius r, so name this region K. Suppose that $\gamma(K) \cap A_0$ is not empty for some non-trivial $\gamma \in \Gamma$. This means that there is some point $x \in A_0$ such that $d_{\mathbb{H}^n}(x_0, \gamma^{-1}x) \leq r$. We can compute then that

$$d_{\mathbb{H}^n}(x_0, \gamma^{-1}x_0) \leq d_{\mathbb{H}^n}(x_0, \gamma^{-1}x) + d_{\mathbb{H}^n}(\gamma^{-1}x, \gamma^{-1}x_0) \leq r + d_{\mathbb{H}^n}(x_0, x)$$

from the triangle inequality. Because $x \in A_0$, we know that $d_{\mathbb{H}^n}(x_0, x) \leq d_{\mathbb{H}^n}(x_0, \gamma^{-1}x) \leq r$, so we combine these inequalities to get

$$d_{\mathbb{H}^n}(x_0, \gamma^{-1}x) \leq 2r.$$

By the compactness of the sphere, there can only be finitely many neighbors to avoid an accumulation point. □

2.7 Space forms

We now tackle the Riemannian definition of hyperbolic manifolds. The contents and methods of this section will not be utilized in the proofs of Mostow rigidity, and therefore may be skipped for a reader without an interest in Riemannian geometry.

The Riemannian definition of a hyperbolic manifold is given by having constant sectional curvature (CSC[7]) -1 everywhere. We recall the notion of sectional curvature given

[7] Not to be confused with constant scalar curvature, often denoted csc.

before in Proposition 2.1.10, and we verified that indeed $\mathbb{H}^n$ is a CSC manifold with sectional curvature -1 everywhere. We will compute that such a manifold can be rescaled such that its sectional curvature can be assumed to be $-1, 0, 1$.

Theorem 2.7.1 (CSC classification)**.** *Let M be a complete and connected manifold of constant sectional curvature K. Such a manifold is often called a space form. The universal cover $\tilde{M}$ is isometric to one of the following, depending on the sign of K:*

(a) $\mathbb{H}^n$ *if* $K = -1$

(b) $\mathbb{R}^n$ *if* $K = 0$

(c) S^n *if* $K = 1$.

Let g be a CSC metric on M. Given a conformal change of the metric $\tilde{g} = e^{2\varphi} g$, similar computations to Proposition 2.1.10 will give the formula

$$\tilde{R}_{ijk\ell} = e^{2\varphi} R_{ijk\ell} - e^{2\varphi} \left(g_{ik} T_{j\ell} + g_{j\ell} T_{ik} - g_{i\ell} T_{jk} - g_{jk} T_{i\ell} \right) \tag{2.19}$$

for $T_{ij} = \nabla_i \nabla_j \varphi - \nabla_i \varphi \nabla_j \varphi + \frac{1}{2} |d\varphi|^2 g_{ij}$. Notably, if $\varphi \equiv c$ is constant, then $T_{ij} \equiv 0$ vanishes. In this case, $\tilde{R}$ is simply the scaled R. This tells us that the computation for the sectional curvature conformally scaled by a constant is

$$\tilde{K}_{ij} = \frac{\tilde{R}_{ijij}}{\tilde{g}_{ii} \tilde{g}_{jj} - \tilde{g}_{ij}^2} = e^{-2c} K_{ij} \tag{2.20}$$

in local coordinate coordinates using a frame of the tangent space. Therefore, we can always re-parameterize by setting $e^{-2c} = |K^{-1}|$ to reduce to one of the three cases above.

The universal $\pi : \tilde{M} \to M$ inherits a metric of constant curvature by pulling back the metric on M to be either flat or have constant sectional curvature ± 1.

An important theorem used to prove the classification of simply connected space forms is due to Cartan, informally stating that locally the metric can be recovered from the curvature. For the setup, let $f : M \to N$ be a smooth map between manifolds of the same dimension. Let $p \in M$ and $q \in N$ and $i : T_p M \to T_q N$ be a linear isometry between the tangent spaces at these points. Define $p \in V \subset M$ to be a normal neighborhood of p such that $\exp_q$ is defined at $i \circ \exp_p^{-1}(V)$. Define $f : V \to N$ as $f(v) = \exp_q \circ\, i \circ \exp_p^{-1}(v)$.[8] For all $v \in V$, there exists a unique normalized geodesic $\gamma : [0, t] \to V$ from p to v. Let P_t be the parallel transport from p to v along γ and $\tilde{P}_t$ is the parallel transport on N along $\tilde{\gamma}$ the geodesic defined by $\tilde{\gamma}(0) = q$ and $\frac{d}{dt}\tilde{\gamma}\Big|_{t=0} = i(\dot{\gamma}(0))$. Define $\phi_t : T_v M \to T_{f(v)} N$ by $\phi_t(v) = \tilde{P}_t \circ\ i \circ P_t^{-1}(v)$. Denote R and $\tilde{R}$ to be the Riemann curvatures of M and N, respectively.

[8]The map $\exp(-)$ is the exponential map which takes in a tangent vector and flows along the unique geodesic emanating from that tangent vector one unit of time. The Hopf-Rinow theorem has a third component saying that this map is defined on the entire tangent space for a complete manifold, so this is well-defined in this context.

Theorem 2.7.2 (Cartan)**.** *If for all $v \in V$ and any $V, W, P, Q \in T_vM$, the equality*

$$R_{ijk\ell}V^iW^jP^kQ^\ell = \tilde{R}_{ijk\ell}\phi_t(V)^i\phi_t(W)^j\phi_t(P)^k\phi_t(Q)^\ell$$

holds, then $f : V \to N$ is a local isometry onto its image and $df_p = i$.

Proof. Consider some $v \in V$, and let $\gamma : [0, T] \to M$ be the normalized geodesic adjoining p to v. For a tangent vector $\mathfrak{v} \in T_vM$, let J be the Jacobi field along γ with initial condition that $J(0) = 0$ and $J(T) = \mathfrak{v}$. Let e_i be an orthonormal frame of T_vM with $e_1 = \dot{\gamma}(0)$ and $e_i(t)$ be the parallel transport along γ to $\gamma(t)$. Locally, $J(t)$ is expressible using this basis and we can label the coefficients by functions $y^i(t)$, meaning $J(t) = y^i(t)e_i(t)$. The Jacobi equation is given by

$$\ddot{y}^k(t) + |\dot{\gamma}|^2 y^j(t)R_{1j1k} = 0 \tag{2.21}$$

using e_i as the coordinate system. Note, the above gives separate equations for each k. It is not summed because the upper and lower indices of k do not appear in a product. Define $\tilde{\gamma}$ to be the normalized geodesic originating at q with $\dot{\tilde{\gamma}}(0) = i(\dot{\gamma}(0))$ and $\tilde{J}(t)$ be the field given by $\phi_t(J(t))$ for $t \in [0, T]$. Similarly, let $\tilde{e}_i(t) = \phi_t(e_i(t))$. Since ϕ_t is linear, we have a corresponding equation $\tilde{J}(t) = y^i(t)\tilde{e}_i(t)$.

The hypothesis that contracting any four vector fields using the curvature is invariant under passing the vector fields through ϕ_t can be applied to say that $R_{1j1i} = \tilde{R}_{1j1r}$ by letting V, W, P, Q be e_1, e_j, e_1, e_r respectively. The Jacobi equation 2.21 can be expressed then as

$$\ddot{y}^k(t) + |\dot{\gamma}|^2 y^j(t)\tilde{R}_{1j1k} = 0, \tag{2.22}$$

demonstrating that $\tilde{J}$ is a Jacobi field along $\tilde{\gamma}$ with $\tilde{J}(0) = 0$. Since parallel transport is an isometry, $|\tilde{J}(T)| = |J(T)|$.

At the initial point, $\dot{\tilde{J}}(0) = i(\dot{J}(0))$ by definition. Since both J and $\tilde{J}$ are Jacobi fields that vanish at 0, they satisfy

$$J(t) = (d\exp_p)_{t\dot{\gamma}(0)}(t\dot{J}(0)), \quad \tilde{J}(t) = (d\exp_q)_{t\dot{\tilde{\gamma}}(0)}(t\dot{\tilde{J}}(0)). \tag{2.23}$$

This can be rearranged to yield the equality

$$\tilde{J}(T) = (d\exp_q)_{T\dot{\tilde{\gamma}}(0)}Ti(\dot{\tilde{J}}(0)) = (d\exp_q)_{T\dot{\tilde{\gamma}}(0)} \circ\ i \circ ((d\exp_p)_{T\dot{\gamma}(0)})^{-1}(J(T)) = df_v(J(T)). \tag{2.24}$$

Showing that $\tilde{J}(T) = df_v(\mathfrak{v}) = df_v(J(T))$ finishes the proof. □

This theorem will be applied to spaces of constant curvature to demonstrate isometries using Corollary 2.7.4 mainly.

Corollary 2.7.3. *Let M and N be manifolds of the same dimension and same constant curvature. Let $p \in M$ and $q \in N$. For e_i an orthonormal basis of T_pM and f_i an orthonormal basis of T_qN, there exists a $p \in V \subset M$ normal neighborhood V of p and corresponding $q \in U \subset N$ and an isometry $F : V \to U$ such that $dF_p(e_i) = f_i$.*

Proof. Take i as mapping $e_i \mapsto f_i$ to fit into the criteria of Theorem 2.7.2 and the curvature criteria is immediately met. $\square$

Letting $N = M$ in the above gives the following:

Corollary 2.7.4. *Let M be a space of constant curvature. For $p, q \in M$ distinct points, let e_i and f_i be orthonormal bases of T_pM and T_qM, respectively. There exist neighborhoods $p \in V$ and $q \in U$ and an isometry $F : V \to U$ such that $dF_p(e_i) = f_i$.*

These results will be used to prove Theorem 2.7.1

Proof of the CSC Theorem 2.7.1. The flat and negatively curved manifolds M with metric g of constant sectional curvature 0 or -1 can be proven together. Let $\pi : \tilde{M} \to M$ be the universal cover with induced metric $\pi^* g$. Let Δ denote $\mathbb{R}^n$ or $\mathbb{H}^n$ for the two cases and this method will handle both simultaneously. Let $p \in \Delta$ and $q \in \tilde{M}$ and fix a linear isometry $i : T_p(\Delta) \to T_q(\tilde{M})$. Consider the map

$$f : \exp_q \circ\, i \circ \exp_p^{-1} : \Delta \to \tilde{M}.$$

Since Δ is complete and simply connected, (as known for $\mathbb{R}^n$ and proven for $\mathbb{H}^n$ in Proposition 2.2.3), with non-positive sectional curvature, this map f is well-defined. Theorem 2.7.2 states that f is a local isometry. Showing further that f is a covering space will prove that it is a global diffeomorphism as $\tilde{M}$ is complete and simply connected, so any connected covering space is a diffeomorphism. This is proven via Lemma 2.7.5.

For positive curvature, the above argument does not work exactly because the map f would not be well-defined on all of S^n. We can rectify this by removing a point to define it on the punctured sphere and repeating this using a different point. Then Lemma 2.7.6 can be applied, that if two isometries agree on a point and their derivatives agree, they are the same. For the final case of positive sectional curvature, it must be shown that there exists some $f : S^n \to \tilde{M}$ which is a diffeomorphism. Fix $p \in S^n$ and $q \in \tilde{M}$ and a linear isometry $i : T_p S^n \to T_q \tilde{M}$. For $-p$ the antipode of p, define the map

$$f : \exp_q \circ\, i \circ \exp_p^{-1} : S^n \setminus \{-p\} \to \tilde{M},$$

and from Theorem 2.7.2, f is a local isometry. For some p' in $S^n \setminus \{p, -p\}$, define $q' = f(p')$ and $i' = df_{p'}$. Define f' as

$$f' = \exp_{q'} \circ\, i' \circ \exp_{p'}^{-1} : S^n \setminus \{-p'\} \to \tilde{M}.$$

For $n > 1$, $S^n \setminus \{p_1, \ldots, p_n\} \cong \mathbb{R}^n \setminus \{p_1, \ldots, p_{n-1}\}$ is connected, so notably $S^n \setminus \{-p, -p'\}$ is connected. The key point here is that $f(p') = q' = f'(p')$ and their derivatives $df_{p'} = i' = df'_{p'}$, so Lemma 2.7.6 gives that that $f = f'$ on $S^n \setminus \{-p, -p'\}$. Therefore, we can define $F : S^n \to \tilde{M}$ by f on $S^n \setminus \{-p\}$ and by f' on $S^n \setminus \{-p'\}$, which is well-defined since f and f' agree on the overlap. F is a local isometry as it is locally defined by isometries and since S^n is compact, this is a covering space since $\tilde{M}$ is simply connected. Therefore, F is a diffeomorphism completing the proof. $\square$

Lemma 2.7.5. *Let M be a complete Riemannian manifold with metric g. Suppose $f : M \to N$ is a local isometry between manifolds of the same dimension. If for all $p \in M$ and all $v \in T_pM$ the inequality $|df_p(v)|_g \geq |v|_g$ holds, then f is a covering map.*

Proof. In the category of manifolds, it is sufficient to show that a map satisfies the unique path lifting property to show that it is a covering space. To demonstrate this, a unique lift $\tilde{\gamma} : [0,1] \to M$ must be given for any $\gamma : [0,1] \to N$ such that $f \circ \tilde{\gamma} = \gamma$ and starting point $\tilde{\gamma}(0) = c$ for any c such that $f(c) = \gamma(0)$.

To produce this, fix some starting point c such that $f(c) = \gamma(0)$. Because f is a local isometry, for some $\varepsilon > 0$ there exists a short time existence of a $\tilde{\gamma} : [0, \varepsilon)$ such that $f \circ \tilde{\gamma} = \gamma|_{[0,\varepsilon)}$. The idea is to re-center the analysis at $\tilde{\gamma}(\varepsilon)$ to extend this curve and use the method of continuity to cover all the points in $[0,1]$ on γ. Consider the set of achievable values

$$A = \{t \in [0,1] : \tilde{\gamma} \text{ can be extended to a lift over } \gamma|_{[0,t]} \text{ with } \tilde{\gamma}(0) = c\}.$$

By definition, $0 \in A$ by fixing a preimage c of $\gamma(0)$. This is guaranteed by the fact that f is a local isometry.

The above argument shows that A is open. Therefore, to show that $A = [0,1]$, since it is non-empty, it remains to be shown that A is closed. Suppose $[0, t_0) \subset A$ and consider an increasing sequence $t_n \to t_0$. Since γ is continuous and M is complete, the set $\tilde{\gamma}(t_n)$ has an accumulation point and is therefore contained in some compact set $K \subset M$. Suppose this were not true, then the computation of the length of the curve from c to $\tilde{\gamma}(t_n)$ would diverge as

$$\begin{aligned} L(\gamma|_{[0,t_n]}) &= \int_0^{t_n} \left|\frac{d\gamma}{dt}\right| dt \\ &= \int_0^{t_n} \left|df_{\tilde{\gamma}(t)}\left(\frac{d\tilde{\gamma}}{dt}\right)\right| dt \\ &\geq \int_0^{t_n} \left|\frac{d\tilde{\gamma}}{dt}\right| dt \\ &\geq d_{g_M}(\tilde{\gamma}(0), \tilde{\gamma}(t_n)), \end{aligned} \tag{2.25}$$

which would imply that the distance between $\tilde{\gamma}(0)$ and $\tilde{\gamma}(t_n)$ in M is unbounded. The Hopf-Rinow theorem says that completeness implies that closed and bounded sets are metrically and geodesically complete, so if this is unbounded, then it violates that M is complete. Therefore, $\{\tilde{\gamma}(t_j)\} \subset K \Subset M$ the points are contained in some compact subset K. By compactness, this has an accumulation point $r \in M$. Let $V \ni r$ be a neighborhood of M such that f is a diffeomorphism between V and its image. It must be that $\gamma(t_0) \in f(V)$, so by continuity, γ must map some interval $I = [a,b] \subset [0,1]$ into $f(V)$ with $a < t_0 < b$. For $n \gg 0$ large enough, $\tilde{\gamma}(t_n)$ is contained in V, and we take the lifting given by f as a local diffeomorphism of $\gamma(I)$ through the point r in M. This lift will agree with $\tilde{\gamma}$ on $[0, t_n) \cap I$ since f is locally a diffeomorphism in this region. This demonstrates an extension of $\tilde{\gamma}$ to t_0 showing A is closed. The conclusion now follows since the path lifting property is now verified, so M is indeed a covering space. $\square$

Lemma 2.7.6. *Let $f, g : M \to N$ be two local isometries between two connected manifolds of the same dimension. If $f(p) = g(p)$ and $df_p = dg_p$ for some point p, then $f = g$ globally.*

Proof. Since f, g are local isometries, df_p and dg_p are non-trivial and there exists a normal neighborhood $p \in V \subset M$ such that $f|_V$ and $g|_V$ are diffeomorphisms. Define $\phi = f^{-1} \circ g : V \to V$ which has the property that $\phi(p) = p$ and $d\phi_p = \mathrm{Id}$. For any $v \in V$, there exists a unique vector $\mathfrak{v} \in T_pM$ which exponentiated is v, $\exp_p(\mathfrak{v}) = v$. Since $d\phi_p = \mathrm{Id}$, $\phi(v) = v$ as well for all such v. Therefore, $f|_V = g|_V$. This procedure can now be re-centered at any $v \in V$ to extend to a larger neighborhood. If M were compact, this would complete the proof since it is connected.

For M not connected, an argument must be made to extend this along arbitrary paths from p to an arbitrary point $q \in M$ using the method of continuity. Let $\gamma : [0, 1] \to M$ be a path from p to q, guaranteed by the connectedness of M. Define the set

$$A = \{t \in [0, 1] : f(\gamma(t)) = g(\gamma(t)),\ df_{\gamma(t)} = dg_{\gamma(t)}\}$$

of times along γ where f and g and their derivatives agree. $0 \in A$ by assumption. The above argument shows that A is open, and in particular $\sup A > 0$. A is also closed since the first property is closed by continuity and the second by continuity of the derivative. Therefore, $\sup A = 1$ and $f = g$ can be globally extended, completing the proof. $\square$

The takeaway from the CSC theorem is that any hyperbolic manifold M is expressible as $\mathbb{H}^n/\Gamma$ for Γ a discrete subgroup of the isometry group of $\mathrm{Isom}(\mathbb{H}^n) = \mathrm{O}^+(n, 1)$.

2.8 Volume

In this section, we will focus on concrete computations of the volume hyperbolic simplices, the higher-dimensional analogs of triangles and tetrahedra. The standard simplex of dimension k is given in $\mathbb{R}^{k+1}$ as the convex hull of the points $e_i = (0, \ldots, 0, 1, 0, \ldots, 0)$ with the 1 being the in i^{th} slot. This can be expressed as

$$\Delta^k = \{(t_0, \ldots, t_k) \in \mathbb{R}^{k+1} : t_i \geq 0, \sum t_i = 1\}.$$

Any polyhedra can be subdivided into simplices, so understanding these volumes gives a complete method of volume calculation. We will use the word volume to mean the volume in any number of dimensions. In particular, hyperbolic volume in dimension two is the area, and we use these terms interchangeably. Those of most interest will be the ideal simplices whose edges are at infinity. A fundamental notion in hyperbolic geometry that separates it from plane geometry is that the volumes of simplices are determined uniquely by their shape; there is no concept of *scaling* the shape. While in flat geometry, we have a notion of similar triangles, two triangles that have all the same angles, but not the same area.

This cannot occur in the hyperbolic setting as stretching will distort the angles; there is no way to *zoom in* on hyperbolic space. In a hyperbolic triangle or tetrahedron, the volume is uniquely determined by its angles or dihedral angles, respectively.

2.8.1 Two dimensions

Theorem 2.8.1 (Angle defect)**.** *The area of a hyperbolic triangle with angles α, β, and γ is the angle defect $\pi - \alpha - \beta - \gamma$. Notably, the area of any ideal triangle is π.*

Proof. Using the triply-transitive property of $\mathrm{PSL}(2, \mathbb{R})$, we can assume the ideal triangle has vertices at ± 1 and ∞, so the area can be computed by integrating

$$K = \int_{-1}^{1} \int_{\sqrt{1-x^2}}^{\infty} \frac{1}{y^2}\, dy dx = \int_{-1}^{1} \frac{1}{\sqrt{1-x^2}}\, dx = \pi. \tag{2.26}$$

Assume that a triangle has a single vertex at ∞, so it has two vertical sides and a semi-circular arc of radius 1 connecting them . Assume that these vertical sides have x coordinate given by $a < b$. Therefore, the area looks like the above integral, but over the region $x \in [a, b]$. In this setting, we get a similar integral, but the domain of x changes, so we can compute

$$K = \int_{a}^{b} \int_{\sqrt{1-x^2}}^{\infty} \frac{1}{y^2}\, dy dx = \int_{a}^{b} \frac{1}{\sqrt{1-x^2}}\, dx,$$

and here a and b can be given by their angles with which they meet the vertical lines, which we will call α and β (see Figure 2.12). These are the same angles that correspond to the (minimal) angle between the radius connecting the vertices to the origin and the x-axis. This is the minimum of the argument and π minus the argument. We can set $x = \cos(\theta)$, and then this area is computed as

$$K = \int_{\pi-\alpha}^{\beta} -\frac{\sin(\theta)}{\sqrt{1-\cos^2(\theta)}}\, d\theta = \pi - \alpha - \beta$$

as expected.

The final case is given a triangle with vertices at interior points of $\mathbb{H}^2$, we can compute its area as a difference. Take a triangle and extend one of its sides to an infinite geodesic, and we see the difference of areas depicted below in Figure 2.13. $\square$

2.8.2 Three dimensions

Ideal tetrahedra are parameterized by Euclidean triangles up to similarity. Without loss of generality, we can allow one vertex to be at ∞ in $\mathbb{H}^3$. We can then take the link of the vertex $v = \infty$ which is the intersection with a *horocycle*, which will be any plane parallel to $\{x^3 = 0\}$. Therefore, the intersection of this link with the tetrahedron is a Euclidean triangle with angles $\alpha + \beta + \gamma = \pi$, which are the adjacent dihedral angles of the ideal

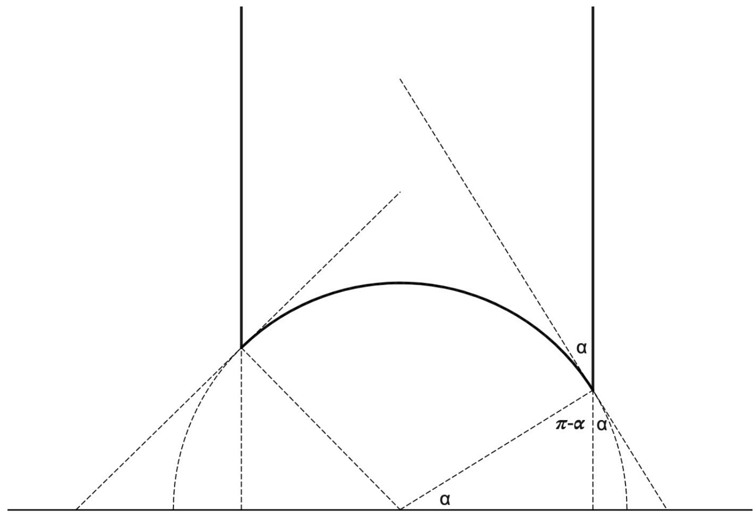

Figure 2.12: The solid black lines form an ideal triangle with a single vertex at infinity in $\mathbb{H}^n$. Let the two vertices be complex numbers of unit length with arguments α and β. The picture shows that the angles made between the tangent lines to the circle and the vertical edges of the hyperbolic triangle are also α and β.

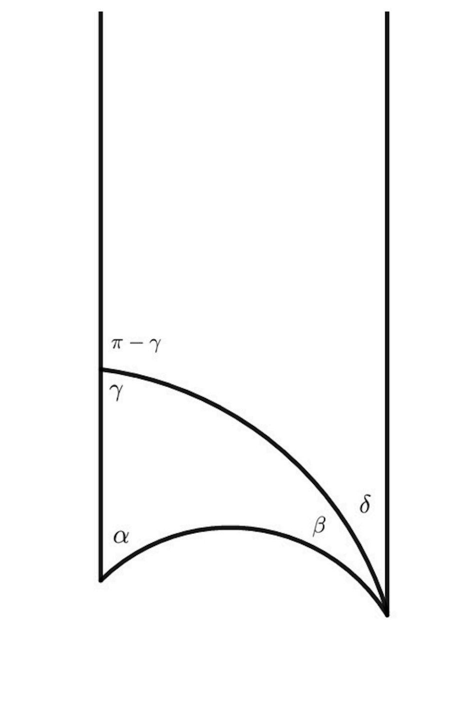

Figure 2.13: Given a triangle with angles α, β, γ, the side between the angles α and β can be extended to infinity. We get two triangles of the form above. Call the unlabeled angle δ. The area of the large infinite triangle is $\pi - \alpha - \beta - \delta$. The area of the smaller infinite triangle is $\gamma - \delta$. Therefore, the area of the finite triangle is their difference, which is $\pi - \alpha - \beta - \gamma$, proving the claim.

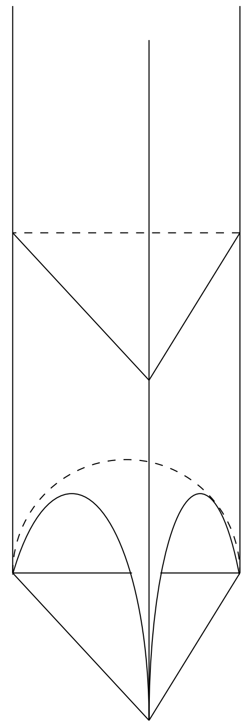

Figure 2.14: We show an ideal tetrahedron in $\mathbb{H}^3$ with one vertex at ∞. The link of the ideal tetrahedron is seen to be Euclidean triangle, showing that triples of adjacent dihedral angles sum to π.

tetrahedron. Here, adjacent dihedral angles are triples of angles that all border a single vertex.

There are four 3-tuples of such dihedral angles. We can label dihedral angles as opposite if they are the pairs of angles between pairs of disjoint vertices. Label the dihedral vertices opposite α, β, γ as α', β', γ'. Therefore, applying the above argument gives equations

$$\begin{aligned} \alpha + \beta + \gamma &= \pi \\ \alpha + \beta' + \gamma' &= \pi \\ \alpha' + \beta' + \gamma &= \pi \\ \alpha' + \beta + \gamma' &= \pi \end{aligned}$$

and we can equate the sum of the first two equations and last two, showing that $\alpha = \alpha'$, so the same follows for the other pairs. Therefore, $\alpha + \beta + \gamma = \pi$ defines the ideal tetrahedron.

Definition 2.8.2 (Lobachevsky function)**.** We define $\Lambda(x)$ to be a periodic function called the *Lobachevsky function*, defined by the integral below and graphed in Figure 2.15.

$$\Lambda(x) = -\int_0^x \log|2\sin t|\, dt$$

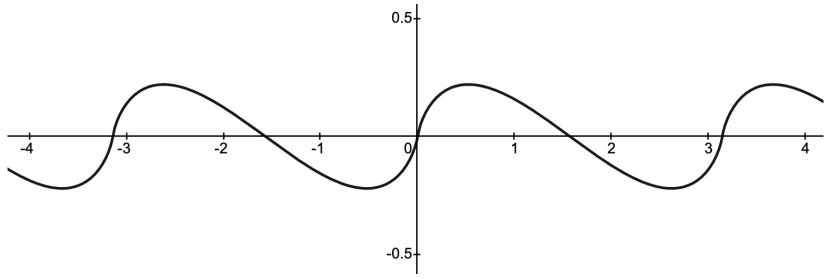

Figure 2.15: The Lobachevsky function is periodic.

shown below in Figure 2.15.

Theorem 2.8.3 (Volume of ideal tetrahedron)**.** *The volume of an ideal hyperbolic tetrahedron with dihedral angles* $\alpha + \beta + \gamma = \pi$ *is given by* $\Lambda(\alpha) + \Lambda(\beta) + \Lambda(\gamma)$*.*

We refer the reader to Ratcliffe [Rat94] for a proof of this theorem. The techniques involved in this proof are not used further in this paper, so it is omitted. The main idea is to do a similar integration as in the two-dimensional case, but the base is a sector of a sphere, so it is more complicated. The main result we will use is that this is maximized for a tetrahedron with all dihedral angles $\frac{\pi}{3}$.

2.8.3 Higher dimensions

The property of the maximal simplex being *regular*, fully symmetric, is true in higher dimensions, but was not known until 1981. This result can generalize Gromov's proof of Mostow rigidity to all dimensions $n \geq 3$; the techniques developed are not utilized elsewhere in this paper, so this is presented in Appendix A. Notably, this will cover the above case in dimension three and prove the necessary result that the ideal regular simplex is of maximal volume.

2.9 Riemann Surfaces

"*The Uniformization Theorem* - the grand-
daddy of all hyperbolic geometry"
-John Hubbard [Hub06]

The rigidity of hyperbolic manifolds is a phenomenon of higher dimensions, and this failure of rigidity for Riemann surfaces is captured, for example, by the moduli space $\mathcal{M}_g$. We have already seen this in Examples 2.5.12 and 2.5.13, which indicated the flexibility to construct Riemann surfaces and the dimension of the moduli space $\mathcal{M}_g$. The uniformization theorem in dimension two is stronger than the classification of higher dimensional space forms, Theorem 2.7.1, because it states that every Riemann surface can be endowed with a CSC metric, while in higher dimensions, not every manifold can be given such a CSC metric.

Theorem 2.9.1 (Uniformization for Riemann Surfaces)**.** *A connected and simply connected Riemann surface is isomorphic to $\mathbb{C}$, $\mathbb{P}^1$ the Riemann sphere, or $\mathbb{B}$, each of which carries a canonical metric of constant curvature.*

These Riemann surfaces cannot be isomorphic to each other. The Riemann sphere is compact. Any holomorphic map $f : \mathbb{C} \to \mathbb{B}$ is bounded and entire, and therefore constant, so there cannot exist an isomorphism between $\mathbb{C}$ and $\mathbb{B}$ (and therefore $\mathbb{H}$). Any open set in $\mathbb{C}$ carries a complex structure induced by $\mathbb{C}$. If $U \subsetneq \mathbb{C}$ is connected and simply connected, uniformization enforces that U must be isomorphic to either $\mathbb{C}$ or $\mathbb{B}$ since it is not compact. If U is bounded, it must be isomorphic to $\mathbb{B}$ from the same argument that any map $\mathbb{C} \to U$ would be constant by Liouville's theorem. It turns out that even if U is not bounded, it is still isomorphic to $\mathbb{B}$ and not $\mathbb{C}$, which combined with the above is the statement of the Riemann mapping theorem.

Theorem 2.9.2 (Riemann mapping theorem)**.** *Let $U \subsetneq \mathbb{C}$ be a connected and simply connected open region. There exists a biholomorphism $f : U \to \mathbb{B}$. Furthermore, we can make this a pointed map so that for any $p \in U$ there exists a biholomorphism $f : (U, p) \to (\mathbb{B}, 0)$.*

For the proof of the Uniformation theorem for Riemann surfaces, we direct the reader to [Hub06] or any comprehensive textbook on complex analysis. Assuming the classification of simply connected space forms Theorem 2.9.1, it would be sufficient to show a single biholomorphic map from U to any bounded region. Since U is not all of $\mathbb{C}$ and simply connected, there must be at least two points $a, b \in \mathbb{C} \setminus U$. Consider the function $f(z) = \frac{z-a}{z-b}$. Translating U such that $0 \notin U$, a choice of a branch of the square root gives two disjoint regions V and $-V$. Suppose for the sake of contradiction they were not disjoint. This would imply there exists some $\zeta \in V \cap -V$. V and $-V$ are given by exponentiating the integral $\frac{1}{2}\frac{df'}{df}$. If the branches agreed at a point, this would violate the well-definedness of this integral, so they must be disjoint. Explicitly, we can take $\log(U)$ and know that this is a one-sided inverse to exp, meaning $\exp(\log(z)) = z$ for all $z \in U$. To compute any value, it is given by $\log(\zeta) = \log(z_0) + \int_{z_0}^{\zeta} \frac{dz}{z}$. By hypothesis, U is simply connected, so this integral is well-defined by just the endpoints. The square root is defined as $\exp(\log(z)/2)$. Pick a point $p \in -V$ and the function $\frac{1}{\sqrt{z-p}}$ is bounded and biholomorphic.

Applying the uniformization theorem above would prove the Riemann mapping theorem 2.9.2, although it is a bit overkill and can be done using the Arzela-Ascoli theorem and the setup above.

Proof of Riemann mapping theorem 2.9.2. Consider all univalent (holomorphic and injective) functions $f : (U, p) \to (\mathbb{B}, 0)$ such that $|f'(p)| > 0$. The above discussion shows that this class of functions is non-empty. We can scale the function f so that $f'(p)$ is real by multiplication by $e^{-i\theta}$ for $\theta = \arg(f'(p))$. The claim is that the function, normalized to have $f'(p)$ real, with the maximum value of $f'(p)$ is indeed a Riemann mapping.

It is a priori unclear that this is well-defined, but follows since this family of functions is a normal family. The derivative of f can be computed using the Cauchy integral formula

$$f'(z) = \frac{1}{2\pi i} \int_{\partial B(z,r)} \frac{f(\zeta)}{(z-\zeta)^2} d\zeta \tag{2.27}$$

and this gives an estimate on the derivative of f as $\frac{1}{r}$ times the maximum of f in the ball. Looking at a small disk around z, the largest the radius could be is the distance to the boundary of U. In this setting, $|f| < 1$, since the image lies in $\mathbb{B}$ and the estimate gives

$$|f'(z)| \leq \frac{1}{d(z, \partial U)}. \tag{2.28}$$

This equations tells us that the derivatives of any univalent function into the disk are bounded, which gives equicontinuity of the functions by the mean value theorem. Applying the Arezela-Ascoli theorem states that any sequence of such functions f_n has a subsequence converging uniformly on compact sets. Furthermore, the derivatives also converge on compact sets as seen by Cauchy's theorem 2.27 and therefore, the limit is indeed holomorphic.

Labeling this limit point f, it must be shown that this is a univalent function with image in the unit disk as well. Suppose otherwise; let $a \neq b \in U$ be two points such that $f(a) = f(b)$. Consider a ball $B(a, \varepsilon)$ such that $\varepsilon < |a - b|$ and the boundary $\partial(a, \varepsilon)$ has no zeros of $f(z) - f(b)$. Since f is holomorphic, this is possible as the zeros are discrete, so their radii from a are discrete values as well. The argument principle states that the integral

$$\frac{1}{2\pi i} \int_{\partial B(a,\varepsilon)} d(\log(f(z) - f(b))) dz \in \mathbb{Z} \tag{2.29}$$

counts the number of values of $f(b)$ contained in $B(a, \varepsilon)$, (minus the number of poles, but there are none). Since $f(a) = f(b)$, this value must be at least 1. However, this function takes only discrete values, so it must be the same when passed under limits of $f_n \to f$ that were univalent functions in the domain analyzed. However, these f_n are injective, so they must have this integral exactly 1, and thus so does f, so f too must be injective. This would fail if f were constant, but that will not matter since the choice of f will be the one with maximal derivative, (and we showed that there exists functions with non-zero derivative at p), so f will not be constant, and thus injective.

Consider a sequence of functions $f_n \to f$ that achieves this maximum. The claim is that f is surjective onto $\mathbb{B}$. Suppose for the sake of contradiction that this were not true. Let a be a point missed by f, $a \in \mathbb{B} \setminus \operatorname{Im}(f)$. The points

$$w_1 = \frac{z-a}{\bar{a}z - 1} \quad \rightsquigarrow \quad w_2 = \sqrt{w_1} \quad \rightsquigarrow \quad b = \sqrt{a} \quad \rightsquigarrow \quad w_3 = \frac{w_2 - b}{\bar{b}w_2 - 1} \tag{2.30}$$

are a composition of functions that does as follows: Firstly, w_1 takes the point a to 0 via a Möbius transformation. Then w_2 takes the square root with a chosen branch. Since 0 was not in the image of w_1, this branch is well-defined as detailed above. Then w_3 undoes this and moves the image back to the unit disk via another Möbius transformation. The idea here is that this composition of functions increased the derivative. Therefore, f could not have been the function with maximal derivative completing the theorem.

The fact that the derivative here increased is either a consequence of the Schwartz lemma, or can be directly computed using the fact that a Möbius transformation $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ has derivative $A' = \dfrac{ad - bc}{(cz + d)^2}$. The composition of all the functions above is a map $F : \mathbb{B} \to \mathbb{B}$, and therefore the Schwartz lemma states that $|F'(0)| < 1$, since it is not a rotation. The function $g = F \circ f$ is such that $g'(0) = F'(f(0))f'(0)$, implying that $f'(0) < g'(0)$, a contradiction. The computation states

$$
\begin{aligned}
g'(0) &= \frac{dw_3}{dw_2}\Big|_{w_2=b} \frac{dw_2}{dw_1}\Big|_{w_1=a} \frac{dw_1}{dz}\Big|_{z=0} f'(0) \\
&= \frac{1}{|b|^2 - 1} \frac{1}{2b}(|a|^2 - 1) f'(0) \\
&= \frac{|b|^2 + 1}{2b} f'(0) \qquad\qquad \text{since } |a|^2 = |b|^4.
\end{aligned} \tag{2.31}
$$

Since $|b| < 1$, this tells us that $g'(0) > f'(0)$. □

2.9.1 Flexibility of Riemann surfaces

The uniformization theorem for Riemann surfaces helps classify them and gives the special result that any Riemann surface carries a canonical metric of constant curvature. Most Riemann surfaces are hyperbolic. For example, in the compact setting, the classification of surfaces says that all compact Riemann surfaces are diffeomorphic to a sphere, torus, or higher genus surface diffeomorphic to a connected sum of tori. These have metrics of constant curvature that are positive, zero, and negative, respectively. This can be seen by the Gauss-Bonnet formula and their Euler characteristics. A genus g Riemann surface Σ_g has Euler characteristic $2 - 2g$, which is positive for $g = 0$, zero for $g = 1$, and negative otherwise.

Example 2.9.3 (Metrics of constant curvature on an annulus). For the non-compact case, annuli give a counter-example to rigidity. Consider $\mathbb{C}^*$ with fundamental group $\mathbb{Z}$. Let's consider how to endow it with the three possible geometries in the most obvious manners. It could be considered as having a round metric $\mathbb{C}^* = \mathbb{P}^1 \setminus \{0, \infty\}$ with the induced round metric as a submanifold of S^2. It could also be endowed with a flat metric similarly as a subset of $\mathbb{C}$. Lastly, it can be endowed with a a metric of negative curvature by taking the point $re^{i\theta}$ and mapping it to $\left(1 - \frac{1}{1+r^2}\right) e^{i\theta}$ and giving it the induced metric from the hyperbolic disk.

These show three ways to give the same underlying structure a metric of positive, zero, and negative constant curvature. The question now is which one of these metrics is *correct*, and why does this not violate the space form classification Theorem 2.7.1? The answer is none of them would be the canonical metric. They do not violate the space form classification Theorem 2.7.1 because none of these metrics are *complete*.

In order to endow a space with the *proper* metric of constant curvature, the map from one of $\mathbb{P}^1, \mathbb{C}$, or $\mathbb{H}$ must be a covering map, which none of the above are. The correct example would be to use the exponential map as the covering map $\exp : \mathbb{C} \to \mathbb{C}^*$, demonstrating that this space has the natural structure of a flat metric. We can see this as $\mathbb{C}/\mathbb{Z}$ with $\mathbb{Z}$ generated by the map $z \mapsto z + 2\pi i$.

This motivates studying other annuli, notably the annulus $\mathbb{B}^*$ and $A(R) = \{z \in \mathbb{C} : 1 < |z| < R\}$. The example of the exponential as a covering motivates how to endow these with metrics of constant curvature as well. The difference between these annuli and $\mathbb{C}^*$ is that their preimages under exponential map are no longer all of $\mathbb{C}$. The preimage of the exponential map of $A(R)$ consists of points z such that $0 < \mathrm{Re}(z) < \log R$, and this is a covering map as the imaginary part differs by $2\pi i$. By applying the Riemann mapping Theorem 2.9.2, this must then be isomorphic to $\mathbb{B}$ and these annuli are endowed with a hyperbolic metric. The last case is $\mathbb{B}^*$, which can be thought of as $A(\infty)$.

Proposition 2.9.4 (Isomorphism classes of annuli). *Every annulus is conformal to $\mathbb{C}^*, \mathbb{B}^*$ or $A(R)$ for some R. Furthermore, $A(r)$ is isomorphic to $A(R)$ if and only if $r = R$.*

In particular, this shows a failure of rigidity as the different annuli above are not conformal to each other, but each has the same fundamental group, namely $\mathbb{Z}$, and $A(R)$ and $\mathbb{B}^*$ are hyperbolic. This shows already a flexibility of hyperbolic structures on the underlying topological space. We can measure this flexibility by the length of the geodesic that generates the fundamental group.

Proof. Let A be an arbitrarily annulus, that is, a region that separates $\mathbb{P}^1$ into two disjoint components. Let 0 be one of these components and ∞ in the other. This considers $A \subset \mathbb{C}$ and the fundamental group of A is generated by a loop that goes around the origin. As above, we take $\tilde{A} = \log(A)$ the region that is the preimage of A under the exponential map, and this is a covering space. The preimages of a certain point differ by $2\pi i n$. This region is simply connected, so it is either all of $\mathbb{C}$, and $A = \mathbb{C}^*$, or via the Riemann mapping theorem, conformal to $\mathbb{H}$. This realizes A as the quotient $\tilde{A}/\mathbb{Z}$ since the deck transformations are the group $\mathbb{Z}$. This can be realized as some $g \in \mathrm{Aut}(\mathbb{H})$ and given then as $\mathbb{H}/\langle g\rangle$ the unit disk modulo the group generated by an infinite cyclic group generated by some automorphism with infinite order. Fix $g(\infty) = \infty$ without loss of generality. If this is the only fixed point, i.e., g is parabolic, then g can be further assumed to be $g(z) = z + 1$. In this case, $A = \mathbb{H}/\langle g\rangle$ is $\mathbb{B}^*$ via the exponential map $\exp(2\pi i z)$. Otherwise, g has a second fixed point, it is hyperbolic, and g can be realized as $g(z) = \lambda z$ for $\lambda > 1$. The exponential map $\exp(2\pi i \log(z/\lambda))$ gives an isomorphism to the annulus $\{\frac{1}{R} < |z| < 1\}$, which composed with the conformal inversion map $\frac{1}{z}$ puts it in the form as above.

To show that $A(r) \cong A(R)$ are conformal equivalents if and only if $r = R$, consider a map $f : A(r) \to A(R)$ that takes the unit circle to itself. The Schwartz reflection principle can then be used to extend to a map $f : B(0, r) \setminus \{0\} \to B(0, R) \setminus \{0\}$ since it can be reflected through the circle to get closer and closer to the origin and the value along each circle of reflection takes values in S^1, satisfying the criteria to apply Schwartz reflection. Similarly, it can be reflected outside the circle of radius r since it takes values in the circle of radius R. This extends $f : \mathbb{C}^* \to \mathbb{C}^*$, so it must be some map $z \mapsto az$ and by the rigidity of conformal functions, $r = R$. $\square$

2.10 Geometric topology

As per Mostow rigidity, there is no moduli of hyperbolic structures on a single manifold, or it is trivially a single point. There is a construction called the *geometric topology* which topologizes all the hyperbolic manifolds of finite volume in dimension three, (specifically including all hyperbolic link and knot complements). It can be constructed in two main ways: the first, which we present here, uses the Gromov-Hausdorff metric, and the second by quantifying a measure of minimal Lipschitz continuity.

2.10.1 Gromov-Hausdorff distance

Definition 2.10.1 (Hausdorff Distance)**.** Given $U, V \subset X$ a metric space, we define the Hausdorff distance between U and V as

$$d_H^X(U,V) = \inf\{\varepsilon : V \subset B_\varepsilon(U), U \subset B_\varepsilon(V)\}. \tag{2.32}$$

Definition 2.10.2 (Gromov-Hausdorff Distance)**.** We define the Gromov-Hausdorff distance between two metric spaces X and Y as

$$d_{GH}(X,Y) = \inf_Z\{d_H^Z(\phi(X), \psi(Y)\} \tag{2.33}$$

for all possible metric spaces Z and all isometric embeddings ϕ and ψ.

Lemma 2.10.3. *Let $Z = X \sqcup Y$ and let d^Z be a metric on Z that agrees with the metrics on X and Y separately. Then,*

$$d_{GH}(X,Y) = \inf_{d^Z}\{d_H^Z(X,Y)\}.$$

Proof. Let d' be the the metric described in the lemma above on $Z = X \sqcup Y$. By definition of Gromov-Hausdorff distance, we know that $d_{GH}(X,Y) \leq d'_H(X,Y)$.

For every $\varepsilon > 0$, there exist isometric embeddings $\phi : X \to Z$ and $\psi : Y \to Z$ such that

$$d_H^Z(\phi(X), \psi(Y)) \leq d_{GH}(X,Y) + \varepsilon.$$

To see this, consider the product metric space $Z \times [0, \varepsilon]$ and the isometric embeddings $\phi_0 = \phi \times \{0\}$ and $\psi_\varepsilon = \psi \times \{\varepsilon\}$. Consider the restriction to $(X \times \{0\}) \cup (Y \times \{\varepsilon\})$ and we have the desired result.

$$
\begin{aligned}
d'_H(X,Y) &\leq d_H^Z(X,Y) \\
&= d_H^{Z\times[0,\varepsilon]}(X,Y) \\
&\leq d_H^{Z\times[0,\varepsilon]}(\phi(X),\psi_\varepsilon(Y)) + d_H^{Z\times[0,\varepsilon]}(\phi_0(X),\psi(Y)) \\
&\leq \varepsilon + d_H^Z(\phi(X),\psi(Y)) \\
&\leq d_{GH}(X,Y) + 2\varepsilon,
\end{aligned}
\tag{2.34}
$$

finishing the proof. □

Gromov-Hausdorff distance is a metric on the space of compact metric spaces up to isometry. Furthermore, it is complete.

For non-compact spaces, the definition of Gromov-Hausdorff convergence is on large compact sets. It will also be necessary to consider pointed maps.

Definition 2.10.4 (ε-Isometry)**.** Let $f : (X,x) \to (Y,y)$ be a pointed map between two metrized spaces. f is an ε-isometry if $|d_X(a,b) - d_Y(f(a),f(b))| < \varepsilon$ and $B_{\varepsilon^{-1}}(y) \subset B_\varepsilon(f(B_{\varepsilon^{-1}}(x)))$. That is, it only changes the distance by at most ε and it takes large balls around x to a region around y of at most ε Hausdorff distance from the corresponding ball around y.

The Gromov-Hausdorff distance between X and Y can be reformulated then as the infimum over all ε of any ε-isometries between X and Y. For non-compact spaces, Gromov-Hausdorff convergence is defined on larger and larger balls around the basepoints.

Definition 2.10.5 (Gromov-Hausdorff Convergence for Non-compact Spaces)**.** Let (X_n, x_n) be a sequence of pointed metric spaces. The pointed metric space (X_∞, x_∞) is the Gromov-Hausdorff limit of the sequence (X_n, x_n) if for any $R > 0$ there exist $R_i > 0$ converging to R and $(\overline{B_{R_i}(x_i)}, x_i) \xrightarrow{GH} (\overline{B_R(x_\infty)}, x_\infty)$.

Definition 2.10.6 (Geometric Topology)**.** Let H be the set of all finite volume hyperbolic 3-manifolds. The geometric topology on H is given by the Gromov-Hausdorff convergence criteria in Definition 2.10.5.

Example 2.10.7 (Convergence to punctured torus)*.* Consider the genus two surface. There is a unique geodesic representing the homology class that loops between the two donut holes. Consider a sequence of genus two surfaces equipped with hyperbolic metrics such that the length of the above geodesic goes to infinity. Furthermore, consider a fixed marking on the left side of the surface. As the distance between the two donut holes gets further and further, the Gromov-Hausdorff limit from the marked point no longer sees the second hole, and the limit is a punctured torus where the geometry near the puncture is a cusp.

The fact that the geometric topology is determined over pointed spaces is an especially important fact, as demonstrated in this example. If the basepoint were to be placed consistently *equidistant* from both holes, the geometry would look like a tube of incredibly high

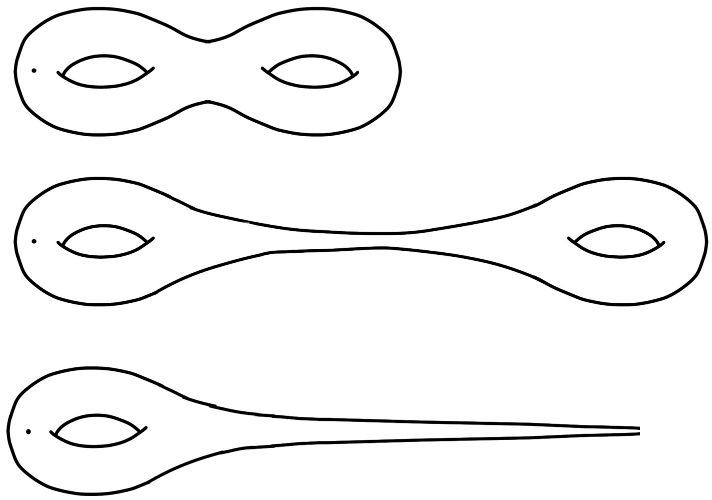

Figure 2.16: As the distance between the two donuts increases, the limit in the geometric topology converges to a punctured torus that has a cusp of finite volume. Figure created by Kalia Firester.

scalar curvature. If the injectivity radius, roughly half the distance around the tube, were to go to 0, then the geometry limit would be a line, collapsing to a single dimension.

The above example indicates that hyperbolic manifolds with a single cusp are limits of compact hyperbolic manifolds. A similar procedure can be done to show that a non-compact hyperbolic with $n+1$ cusps can be obtained as the geometric limit of a hyperbolic manifold with n cusps. This motivates the following theorem detailing that in three dimensions, hyperbolic manifolds can be well-ordered by their volume.

Theorem 2.10.8 (Thurston-Jorgensen)**.** *Consider the volume function on the space of hyperbolic 3-manifolds of finite volume. The image forms a non-discrete set on the real line which is well-ordered and of cardinality ω^ω, for ω the first infinite ordinal. While the volume function is not injective, it is close to injective and is finite-to-one, i.e., there are only finitely many hyperbolic manifolds of a given finite volume.*

What this theorem says is we can order the hyperbolic manifolds according to their volumes. Firstly, there is a manifold of smallest volume x_1. Then, there is a next smallest volume $x_2 > x_1$ and so on. These volumes have an accumulation point $x_1 < x_2 < \cdots < x_\omega$. This number turns out to be the volume of a complete, but non-compact, hyperbolic manifold which has a single cusp. This pattern continues, and for example, $x_{2\omega}$ will be the next smallest non-compact hyperbolic volume. These points eventually accumulate at x_{ω^2}, which continuing the pattern, represents the smallest volume of a hyperbolic manifold with two cusps.

Using this, we can represent $\mathcal{A}_0$ to be the compact hyperbolic manifolds. The image under the volume map of $\mathcal{A}_0$ are the points $x_1, x_2, x_3, \ldots$. We iteratively define $\mathcal{A}_{i+1}$ as the closure of $\mathcal{A}_i$. This can be seen in Figure 2.16, where we see that a single cuspidal torus is the limit of compact hyperbolic manifolds in the geometric topology.

2.10.2 Dehn filling

A Dehn filling, or Dehn surgery, is a method to construct further hyperbolic manifolds from a given finite volume hyperbolic 3-manifold. This procedure shows the abundance of hyperbolic 3-manifolds, as *most* Dehn fillings are hyperbolic manifolds, a theorem due to Thurston. If you remove a solid torus representing the link of a cusp, you can fill it back in non-trivially by twisting it by (p, q), i.e. inducing the map (p, q) as $\mathbb{Z}^2 \to \mathbb{Z}^2$ on the fundamental group of the boundary torus where p and q are coprime. We use the notation (∞, ∞) for the identity in both directions, and a single (∞, q) or (p, ∞) to be the identity in only one direction. Let $M_{(p_1,q_1),\ldots,(p_n,q_n)}$ be the Dehn surgery on M with n boundary components, such that on the i^{th} cusp, we perform Dehn surgery with Dehn invariant (p_i, q_i). Notably, $M_{(\infty,\infty),\ldots,(\infty,\infty)} \cong M$.

Theorem 2.10.9. *Only finitely many $M_{(p_1,q_1),\ldots,(p_n,q_n)}$ are not hyperbolic. Furthermore, as $p_i^2 + q_i^2 \to \infty$ for all i, $M_{(p_1,q_1),\ldots,(p_n,q_n)} \to M$ in the geometric topology.*

For example, the figure eight knot complement, see below in Section 2.11, has only 6 exceptional Dehn surgeries that are not hyperbolic. We refer the reader to Chapter 4 of Thurston [Thu79] and Thurston [Thu82] for more details.

2.11 Knot complements

Thurston's geometrization of knot/link complements states that most such structures admit unique hyperbolic structures of finite volume.

Definition 2.11.1 (Knot/Link Complement)**.** A knot is an embedded $S^1 \hookrightarrow S^3$ considered up to homotopy. That is, two knots $\phi, \psi : S^1 \to S^3$ are considered the same if there exists a homotopy map $F : S^3 \times [0, 1] \to S^3$ such that $F_0 = \text{Id}$ and $F_1 \circ \phi = \psi$. A link is multiple, disjoint, embedded knots $L : \bigsqcup S^1 \hookrightarrow S^3$ with the same condition on equivalence.

Definition 2.11.2 (Classification of Knot/Link Complements)**.** A knot $\phi : S^1 \hookrightarrow S^3$ is called a *torus knot* if there exists an embedding of a torus $\psi : S^1 \times S^1 \to S^3$ such that $\text{Im}(\phi) \subset \text{Im}(\psi)$, i.e. the knot is contained on the surface of an embedded, but not knotted, torus.

Consider $V = S^1 \times D^2$ a solid torus. A knot $\phi : S^1 \hookrightarrow S^3$ has a tubular neighborhood that is a non-trivial (up to homotopy) embedding of $f : V \hookrightarrow S^3$. Suppose there exists a knot $\psi : S^1 \to V$ that is non-trivial meaning it is not contained in an embedded $S^3 \subset V$, so it defines a non-trivial homology class, and is not isotopic to the central core $S^1 \times \{0\}$. The knot ϕ is said to be a *satellite knot* if $\phi = f \circ \psi$, i.e. the knot is knotted on the surface of a torus that itself is knotted in the ambient space.

A knot is said to be *hyperbolic* if the complement $S^3 \setminus \phi(S^1)$ admits a complete finite volume hyperbolic metric.

It is a theorem of Thurston that these are all the possible types of knots and they are disjoint.

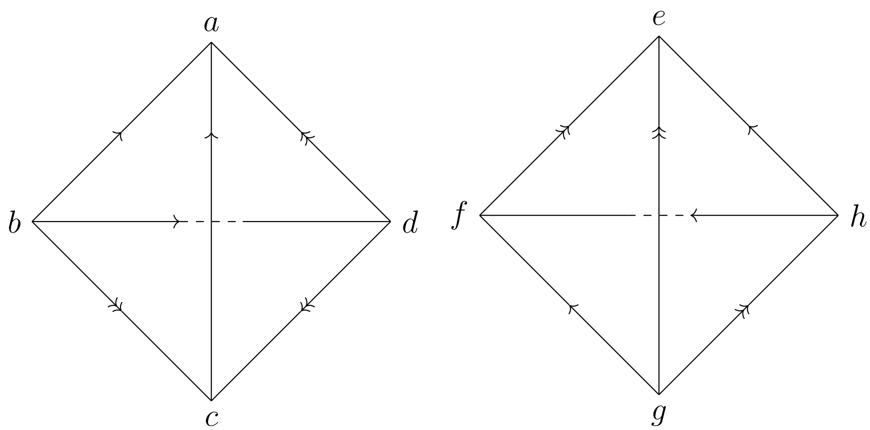

Figure 2.17: The gluing pattern for two ideal tetrahedra to create the figure eight knot complement.

Theorem 2.11.3 (Thurston)**.** *Every knot is either a torus knot, a satellite knot, or hyperbolic. Furthermore, these definitions are disjoint.*

Example 2.11.4 (Figure eight knot complement)*.* The complement of the figure eight knot will be constructed as the gluing of two ideal tetrahedra. Notably, it will be a non-compact hyperbolic manifold with volume $6\Lambda(\pi/3) \approx 2.02988$, which is the smallest volume for a hyperbolic knot/link complement.

Let T and T' be two ideal tetrahedra. Firstly, by viewing these in $\mathbb{H}^3$, this can be viewed as having a vertex at ∞ and three vertices forming an equilateral triangle in the boundary $\mathbb{C}$, points where $x^3 = 0$. We can label them $1, \omega, \omega^2$ for ω a primitive cube root of unity. The geodesics connecting the points form three vertical semicircles and three vertical lines. A large horosphere centered at ∞ is a horizontal plane $x^3 = R$, which intersects this at an equilateral triangle. Therefore, the dihedral angles of such an ideal tetrahedron are all $\pi/3$.

We can label the sides as indicated below to define a unique gluing pattern between the ideal tetrahedra as seen in Figure 2.17. However, viewing this as closed tetrahedra, this is not a manifold as its Euler characteristic is one, as seen by it having a single vertex, two 1-cells, four triangles, and two tetrahedra. In order to define it as a manifold, a point must be removed. The natural point to remove is the specified vertex whose neighborhood is the cone of a torus. The neighborhood of the vertex is a cone over a torus, as seen by removing a small neighborhood and gluing together the triangles at the vertices along the double arrow edges seen in Figure 2.18. The idea that the complement of the point in the figure eight gluing diagram Figure 2.17 from the previous chapter yields a manifold motivates that if the tetrahedra were ideal tetrahedra, that do not include the vertex, the same gluing would yield a hyperbolic manifold. The idea is to find a copy of the figure eight knot (Figure 2.19) along the edges in these tetrahedra.

Claim 2.11.5. The manifold M, as formed by gluing two tetrahedra along the gluing diagram in Figure 2.17 with the vertex removed, is homeomorphic to $S^3 \setminus S^1_8$ where S^1_8 is the figure eight knot (see Figure 2.19).

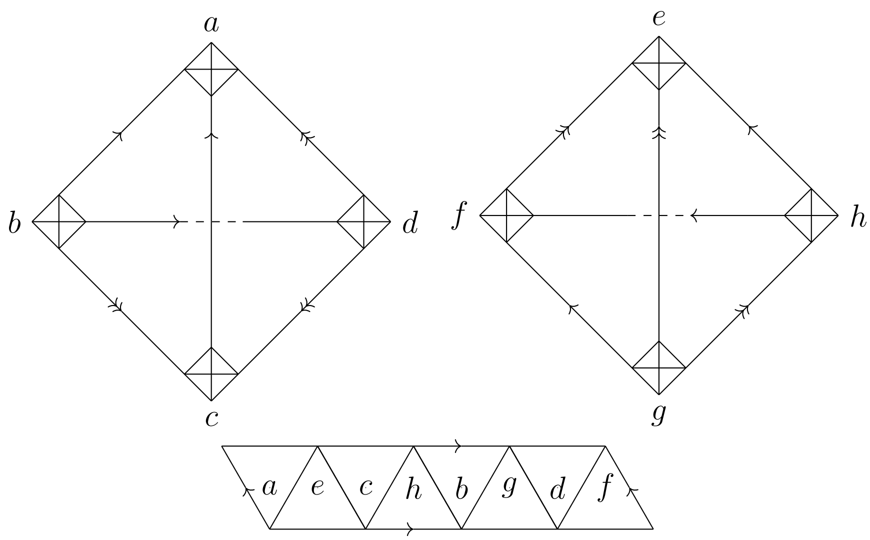

Figure 2.18: A neighborhood of the vertex can be seen to be a cone of a torus. Removing a small triangle parallel to the opposite face near each vertex, (and labeling it as the same letter), the gluing pattern is as follows with opposite sides identified demonstrating this slice is a torus, so as these triangles approach the point, this is a cone over this torus.

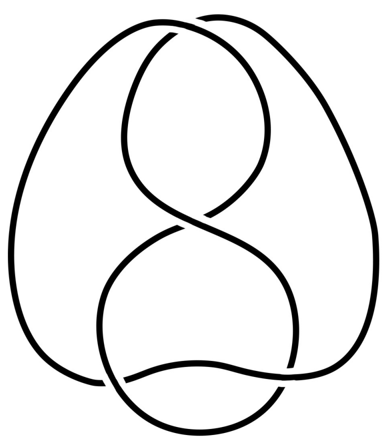

Figure 2.19: The figure eight knot. Figure created by Kalia Firester.

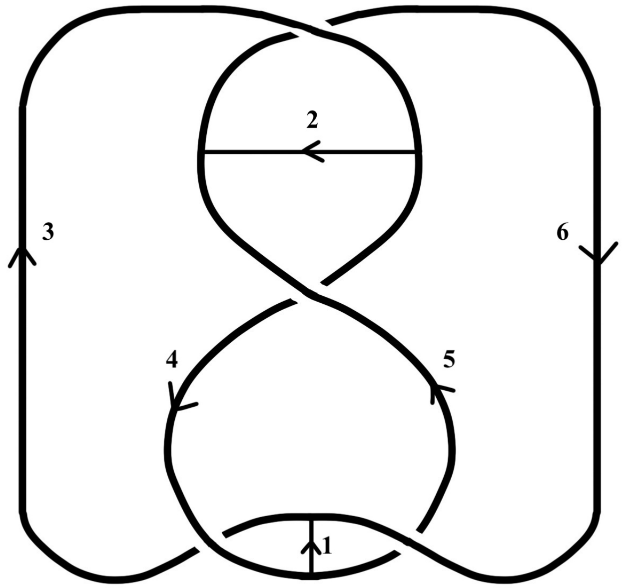

Figure 2.20: The figure eight knot with two edges added to construct the knot complement as a CW complex. Figure created by Kalia Firester.

In Figure 2.20, the knot diagram is shown with added edges (labelled 1 and 2) which will be used to construct this knot as the complement of the ideal hyperbolic tetrahedra glued along the pattern in Figure 2.17. To construct the figure eight knot complement as a gluing of polyhedra, first consider the above knot diagram with the added edges between single DNA loop-looking components. These will be the points to which 2-cells will be glued. A 2-dimensional CW-complex will be constructed utilizing these added edges with four cells given by the regions in the knot diagram as shown below in Figure 2.21. The following lemma is the motivation and main tool of adding in the edges 1 and 2 in Figure 2.20.

Lemma 2.11.6 (Untwisting a DNA loop by adding a base pair). *In the ambient space of S^3, the two diagrams in Figure 2.22 are isotopic.*

Proof. Throughout the process of constructing hyperbolic structures on knot complements, the DNA loop, (with the added edge that can be viewed as crossing the two strands like a single base pair), will need to be deformed from the left image to the right in Figure 2.22. To see this, the base pair edge can be thickened to be a ball and then the two strands can be slid around the ball to untwist the loops. Then, the ball can be shrunk again in the form as desired. This is shown in Figure 2.23. □

Applying Lemma 2.11.6 to the cell complex described in Figure 2.20 with the 2-cells added as shown in Figure 2.21 gives the following 2-dimensional CW-complex in Figure 2.24. Notably, since this is in S^3, the cell B wraps back on itself at infinity. Therefore, the

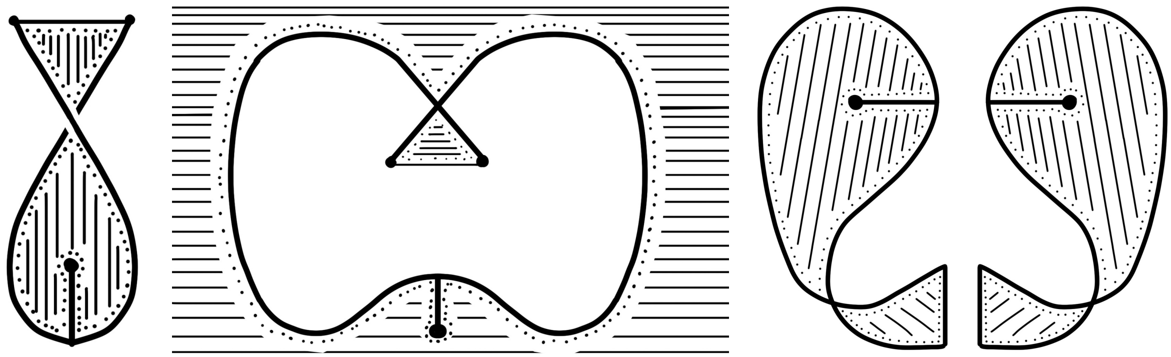

Figure 2.21: The four 2-cells needed to construct the 2-skeleton of the figure eight knot complement structure. The dotted lines show how the gluing map is given for the attached 2-cell. Notably, it goes around the two added edges, which are evidently needed to construct this. Figure created by Kalia Firester.

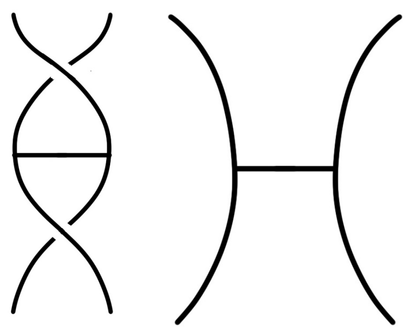

Figure 2.22: The two diagrams are ambiently isotopic in S^3. Adding an edge that resembles a base pair if viewed like a strand of DNA allows the figure to be reformed to the right, and simpler, image. Figure created by Kalia Firester.

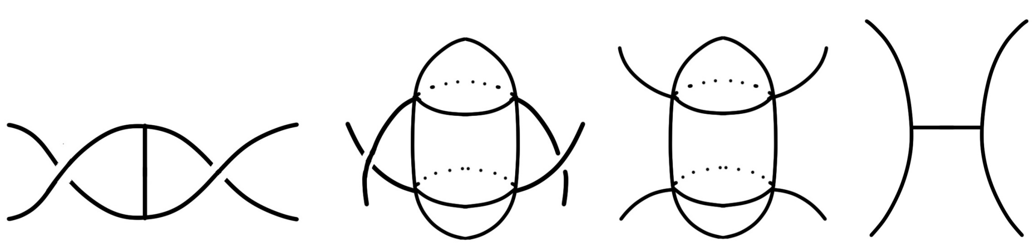

Figure 2.23: Thickening the base pair of a DNA loop to untwist it. The arrows in the second image show how to rotate the strands to untwist them. Figure created by Kalia Firester.

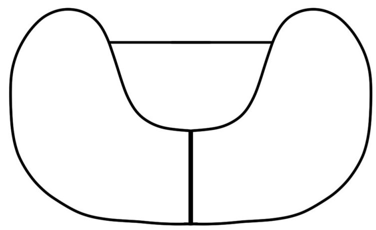

Figure 2.24: The untwisted 2-complex formed from the four cells from Figure 2.21 after untwisting the DNA loops via the added base pair edge. Figure created by Kalia Firester.

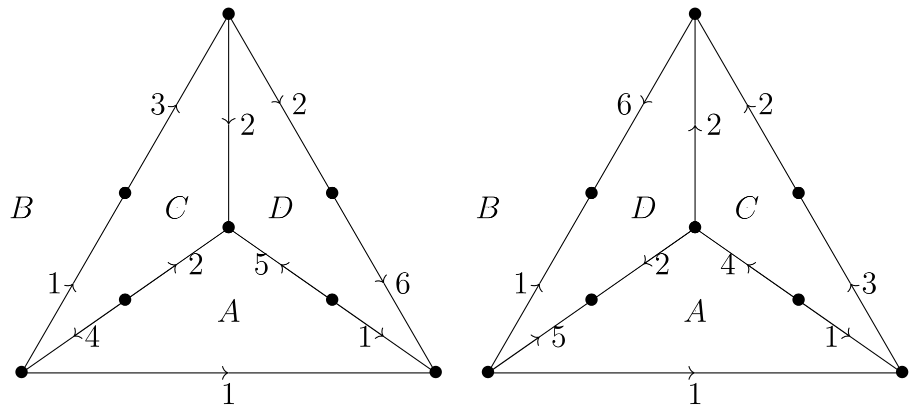

Figure 2.25: The attaching maps of 3-cells creating the inner and outer 3-disks in $S^3 \setminus S^1_8$, by gluing the 3-cells to the faces in Figure 2.24.

complement of this complex is made up of two 3-dimensional balls D^3. Considering this inside S^3 gives that this can be extended through cell attachment of these two 3-cells to be a 3-dimensional CW-complex. These maps are seen below in Figure 2.25. Recall that all the 2-cells were attached to go around the extra base pair added edges, labeled 1 and 2. Figure 2.25 constructs S^3 as a strange CW-complex, but by its construction, it is seen that by removing edges 3,4,5, and 6, the resultant space is $S^3 \setminus S^1_8$ the figure eight knot complement if the vertices were collapsed and removed. This construction, by applying this to ideal tetrahedra, renders the removal of the vertex unnecessary, since ideal polyhedra do not contain their vertices at infinity, and therefore gives this construction a hyperbolic metric as desired. The volume here would be the sum of two ideal polyhedra all with angles $\frac{\pi}{3}$, which would be $6\Lambda\left(\frac{\pi}{3}\right)$ with Λ the Lobachevsky function. This value is roughly 2.02988.

The takeaways of this construction are as follows: Knot complements can be endoeds with a hyperbolic structure if they are given via polyhedral gluing such that they have the proper manifold structure, meaning that the angles need to add up correctly. Otherwise the resultant is rather an orbifold. The DNA untwisting property is necessary, because the resultant CW-structure without the added base pair would not be made of triangles. Rather this would result in a 2-gon which cannot be the face of a polyhedron. Ending up with tetrahedra where the vertices must be removed lends itself to using ideal tetrahedra in hyperbolic space, but is not necessary if the vertices have proper angles. Think of a Euclidean torus as a square with opposite sides identified, and the vertex is kept. The end point was tetrahedra, but if it were not, faces can be further subdivided into triangles to achieve that.

Example 2.11.7 (Alternating knot/link complements). The above methodology can be generalized to all alternating knot/link complements. An alternating diagram means that as you trace across each loop, it alternates going over and under itself. The strategy is similar to the figure eight knot complement structure, so understanding the detailed construction and methodology above will carry most of the techniques and intuition into the more general setting.

Consider L to be an alternating knot/link diagram. Across each crossing, we can add two vertices and a *vertical crossing* edge between them to make the diagram into a combinatorial graph. In addition to such vertices, we will also have edges connecting adjacent links between the upper vertex on one side and the lower vertex on the other side, corresponding to the property that the knot is alternating.

We will define regions in this diagram which will correspond to instructions to construct the hyperbolic structure on the knot/link complement from ideal hyperbolic polyhedra. Consider an n-gon P that corresponds to a region in the knot/link diagram. When considered in the above combinatorial graph, this will correspond to a $2n$-gon as a combinatorial region defined by the interior of a simple cycle in the graph. We call this $2n$-gon $\overline{P}$. We define P' from $\overline{P}$ by removing the edges of P that correspond to knot/link components between adjacent crossings, and define these to be the ideal vertices of P'. Geometrically, the edges of P' are in correspondence with crossings on the boundary of P.

We define ideal polyhedra $B^{\pm}$ by giving their boundaries as $\partial B^{\pm}$, each with a copy of each P' as a face. To glue these ideal polyhedra together, we identify B^+ and B^- along their boundary in P' along the identity map. To complete the picture, we must glue these P' together, which will finish gluing the boundary sides of all the $\partial B^{\pm}$ polyhedra. The above gluing was only partially done on the boundary pieces given by the P' component, which removed some edges.

Consider adjacent regions P and Q in the knot/link diagram. Let e be the edge between them, which goes from an undercrossing e^- to an overcrossing e^+, by the alternating assumption. These crossings of $e^{\pm}$ relate to edges of P^+ and Q^+ in ∂B^+ and ∂B^-. This gives the gluing construction by ∂B^+ in P' and q' along e^- and in ∂B^- in P' and q' along e^+. This gives the final gluing of the boundary pieces of B^+ with all the other regions.

The last components we need to handle are the 2-gons, which will be handled as above in the figure eight knot. As we saw above, these 2-gons can be collapsed to edges and do not alter the topology of the glued picture.

This can be applied to the Borromean rings L and to the Whitehead link, for example, which will correspond to gluing ideal octahedra together, two for the Borromean rings and a single one for the Whitehead link. These link complements will have hyperbolic volumes computable approximately 7.32772 and 3.66386, respectively. These examples can be seen in detail in Thurston [Thu79].

2.12 Algebraic topology preliminaries

One important feature of hyperbolic manifolds is that they are Eilenberg-Maclane spaces. An Eilenberg-Maclane space is a space $K(A, n)$ such that

$$\pi_i(K(A,n)) = \begin{cases} 0, & i \neq n \\ A, & i = n \end{cases} \tag{2.35}$$

has only a single homotopy group, which is A in degree n. Because of the classification of simply connected space forms in Theorem 2.7.1, the universal cover of a connected hyperbolic manifold M is contractible, (in fact diffeomorphic to $\mathbb{R}^n$), so $\pi_i(M) = 0$ for $i \geq 2$. Therefore, $M = K(\pi_1(M), 1)$ is an Eilenberg-Maclane space. Let M and N be two hyperbolic manifolds with the same fundamental group. Let's call this $\pi = \pi_1(M) = \pi_1(N)$.

Proposition 2.12.1. *Let $\rho : \pi \to \pi'$ by any group homomorphism. There exists a unique map up to homotopy $f \in [K(\pi, n), K(\pi', n)]$ such that $f_* : \pi_n(K(\pi, n)) \to \pi_n(K(\pi', n))$ is ρ.*

Corollary 2.12.2. *For M and N hyperbolic manifolds with the same fundamental group π, any automorphism ρ of π can be realized by a map $f : M \to N$ such that $\rho = f_*$ is the induced map on $\pi_1(M) \to \pi_1(N)$.*

One final theorem will we use is the CW-approximation theorem which says that any map can be equivalently, up to homotopy, expressed as a cellular map, i.e., a map that respects a triangulated structure on the manifolds. CW-complexes are constructed by adding simplices one dimension at a time. In the hyperbolic context, these simplices can be given a geometric structure as hyperbolic. If the gluing criteria described in Section 2.5 are met, this forms a well-defined hyperbolic manifold as detailed above.

Theorem 2.12.3 (CW-approximation)**.** *Let $f : X \to Y$ be a continuous map between CW-complexes. The map f is said to be cellular if $f(X^{(n)}) \subset Y^{(n)}$ where $X^{(n)}$ and $Y^{(n)}$ denote the n-skeleton. The CW-approximation theorem says that f is homotopic to a cellular map.*

The proof of this theorem follows by induction on the n-skeleton of X and Y. The 0-simplices can be homotopically moved to the 0-simplices on Y along the 1-simplices. For higher dimension skeletons, we can use compactness of the simplices to similarly perturb an n-simplex off of higher dimensional simplices in Y. We include the algebraic topology constructions and proofs of these theorems in Appendix B and refer the reader to Hatcher [Hat02] for more background in algebraic topology.

3
Mostow's proof of rigidity

3.1 Mostow's proof overview

The first proof of Mostow rigidity shown here will be Mostow's original proof for compact hyperbolic manifolds. It utilizes the compactness to give an initial bound on the modulus of continuity and produce a lift to the universal covers which is a *pseudo-isometry*, i.e., it distorts distance only in a bounded manner. The end goal is to perturb this to an isometry. While isometries take geodesics to geodesics, a pseudo-isometry gets close enough that the image of a geodesic will be a *quasi-geodesic*. We will use the geometry of hyperbolic space to associate to any quasi-geodesic a unique geodesic from which it lies only a bounded distance away. Therefore, our original map will associate geodesics, just like an isometry. We then will use *ergodic theory* to improve the regularity of this map on the boundary to a conformal map on the boundary sphere at infinity. As detailed in Chapter 2 Section 2.3, such a conformal map corresponds to a unique isometry. This will complete Mostow's theorem by deforming the pseudo-isometry to the isometry given by the conformal map on the boundary. In this chapter, we will first go through some introductory definitions and background to define all the terms above and their relationships. With this setup, the proof of rigidity will be quite quick and intuitive.

3.2 Pseudo-isometries and quasi-geodesics

In the context of Mostow rigidity, a given map $f : M_1 \to M_2$ will be a homotopy equivalence with $g : M_2 \to M_1$ its homotopy inverse. Each M_i has its universal cover $\pi_i : \mathbb{H}^n \to M_i$, and the deck transformations are given by $\Gamma_i \subset \mathrm{Isom}^+(\mathbb{H}^n)$. Notably, $\Gamma_1 \cong \Gamma_2$ since f is a homotopy equivalence. Take $\tilde{f}$ to be a lift of f between the universal covers.

Definition 3.2.1 (Pseudo-isometry)**.** A pseudo-isometry is a map $f : X \to Y$ between metric spaces such that there exist constants $C, C' > 0$ satisfying for all $a, b \in X$ the bound

$$C^{-1} d_X(a,b) - C' \leq d_Y(f(a), f(b)) \leq C d_X(a,b). \tag{3.1}$$

This says that f only changes distances in a bounded way with some error. If $C = 1$ and $C' = 0$, then this is an isometry.

A key point of isometries is that they carry geodesics to geodesics. The remarkable feature of hyperbolic geometry is that in this context, this property is almost true for pseudo-isometries. While the image of a geodesic is not a geodesic, it is in a neighborhood of a unique one, so a pseudo-isometry still associates geodesics. To understand this association, we define the notion of a *quasi-geodesic*, and in hyperbolic space, we will see that any quasi-geodesic can be associated to the unique geodesic from which it lies within a bounded distance.

Definition 3.2.2 (Quasi-geodesic)**.** A *quasi-geodesic* is a map $\gamma : I \to X$ where I is an interval (possibly all of $\mathbb{R}$) and X is a metric space, such that γ is a pseudo-isometry.

The first theorem says that the lifts $\tilde{f}$ and $\tilde{g}$ between the universal covers are pseudo-isometries.

Lemma 3.2.3. *$\tilde{f}, \tilde{g} : \mathbb{H}^n \to \mathbb{H}^n$ can be made to be pseudo-isometries by perturbing f and g up to homotopy.*

Proof. By the CW-approximation theorem, we can assume that the maps f and g are cellular maps, (given a cellular structure on M_1 and M_2). Since M_1 and M_2 are compact, there exists a uniform bound $\sup_{a,b \in M_1} d_1(a,b) < C$ and $\sup_{c,d \in M_2} d_2(c,d) < C$. Suppose otherwise, a path from a, b that has infinite length gives infinitely many charts of radius 1 around points on this curve, and including also the complement of the curve to these charts gives an open covering that cannot be reduced to a finite subcover, thereby violating compactness. This tells us that C gives a Lipschitz coefficient for both f and g, which will lift to give that $\tilde{f}$ and $\tilde{g}$ are Lipschitz, that is,

$$d_{\mathbb{H}^n}(a,b) \leq C d_{\mathbb{H}^n}(\tilde{f}(a), \tilde{f}(b)), \qquad d_{\mathbb{H}^n}(c,d) \leq C d_{\mathbb{H}^n}(\tilde{g}(c), \tilde{g}(d)). \tag{3.2}$$

Considering $a, b \in M_1$ and $x, y \in M_2$ to be their images, so $f(a) = x$ and $f(b) = y$. Using similar names for lifts of these points the universal covers, the Lipschitz condition gives bounds

$$d_{\mathbb{H}^n}(\tilde{g} \circ \tilde{f}(x), \tilde{g} \circ \tilde{f}(y)) \leq C d_{\mathbb{H}^n}(\tilde{f}(x), \tilde{f}(y)). \tag{3.3}$$

Since $f \circ g$ and $g \circ f$ are homotopic to the identity, this homotopy gives a path from $g \circ f(x)$ to x. By compactness, this path has length bounded by some constant C'. The same applies to the universal cover maps $\tilde{f}$ and $\tilde{g}$, as they too are homotopy inverses and a path is the same length when lifted by the covering map which is a local isometry. This gives the bound

$$d_{\mathbb{H}^n}(x,y) - 2C' \le d_{\mathbb{H}^n}(\tilde{g}\circ\tilde{f}(x), \tilde{g}\circ\tilde{f}(y)). \tag{3.4}$$

By taking the max of 1 and C, this gives the first inequality in Definition 3.1 to get

$$C^{-1}d_{\mathbb{H}^n}(x,y) - 2C' \le d_{\mathbb{H}^n}(\tilde{g}\circ\tilde{f}(x), \tilde{g}\circ\tilde{f}(y)), \tag{3.5}$$

finishing the proof by replacing C' with $C'/2$. □

While an isometry maps a geodesic to a geodesic, a pseudo-isometry still associates a geodesic to the image of a geodesic, even though it is not exactly the image.

Proposition 3.2.4 (Pseudo-isometries associate geodesics)**.** *Let $F : \mathbb{H}^n \to \mathbb{H}^n$ be a pseudo-isometry with constants C, C' to fit the definition from Equation 3.1 and let γ be a geodesic in $\mathbb{H}$. There is a unique geodesic γ' such that $F(\gamma)$ lies in a bounded set around γ'.*

Proof. Let $B_r(\Sigma)$ for $\Sigma \subset \mathbb{H}^n$ be the set of points of distance less than r to some point in Σ. This language gives a restatement of the proposition in a stronger sense. There exists a constant $r > 0$ such that for every geodesic γ, there exists a unique geodesic γ' such that $F(\gamma) \subset B_r(\gamma')$. Fixing γ, it will be demonstrated that r is not actually a function of γ, and is therefore well-defined. Let $a, b \in \mathbb{H}^n$ be two points and let $[a,b]$ be the geodesic segment adjoining them. The image of this geodesic segment $F([a,b])$ is contained in $B_t([F(a), F(b)])$. To show this, consider first some $t_0 > 0$ such that $\cosh t_0 = C^2 + 1$ (see Lemma 3.2.5 to motivate this choice of constant). Consider some geodesic segment $[x,y]$ that is contained in $[a,b] \cap F^{-1}(B_{t_0}([F(a), F(b)]))$. The distance $d_{\mathbb{H}^n}(F(x), F([a,b]) = d_{\mathbb{H}^n}(F(y), F([a,b])) = t_0$. Let π_γ be the orthogonal projection to the geodesic γ.

Lemma 3.2.5. *Let a, b be two points at distance t from a geodesic γ. Then, $d_{\mathbb{H}^n}(a,b) \ge \cosh(t) \cdot d_{\mathbb{H}^n}(\pi_\gamma(a), \pi_\gamma(b))$.*

Geometrically, this lemma says that the distance between points t away from a geodesic grows exponentially.

Proof. We work in the hyperboloid model of hyperbolic space. Here, we can represent γ as the intersection of $\mathcal{H}^n$ and a linear plane (two dimensional) L through the origin inside of $\mathbb{R}^{n,1}$. We can denote W as the orthogonal space to L with respect to the Minkowski signature $(n,1)$ metric, and let S be the unit sphere in W. We denote the tubular region of distance r from γ by $C_r(\gamma)$ inside $\mathcal{H}^n$.

We define a one-sided inverse map to π_γ which takes a point on γ and a direction from S into $C_r(\gamma)$ by

$$\omega : \gamma \times S \to C_r(\gamma), \qquad \omega : (u,w) \mapsto \cosh(r)u + \sinh(r)w,$$

which is a diffeomorphism from the characterization of geodesics in the Minkowski model (see Remark 2.1.6).

Consider now $u' \in L$ and $w' \in W$, and we can compute the hyperbolic distance composed with this map at the point $(u', w') \in L \times W$ as

$$d_{(u',w')}\omega(u', w') = \cosh(r)u' + \sinh(r)w'$$

and take its norm to get

$$\|d_{(u',w')}\omega(u', w')\| = \cosh^2(r)\|u'\| + \sinh^2(r)\|w'\| \geq \cosh^2(r)\|u'\|.$$

This gives the desired inequality

$$d(\omega(u_1, w_1), \omega(u_2, w_2)) \geq \cosh(r) \cdot d(u_1, u_2)$$

completing the proof.

For another approach, as in many cases in hyperbolic geometry, we can consider the plane going through the geodesic and a third point p, and we can reduce to studying the two-dimensional case. Here, we can consider the simplified question where γ is the imaginary axis and π is the projection map to it. We can rescale our points such that $\pi(p) = i$. Therefore, we want to travel along the geodesic that is the upper half of the unit circle in $\mathbb{C}$. This is parameterized by $t \mapsto \tanh(t) + i \operatorname{sech}(t)$. From Example 2.1.7, we know that the imaginary part of p must be $\operatorname{sech}(r)$. As in that example, the path that takes any point to the imaginary axis is a half-circle centered at 0. Therefore, we know that in the Euclidean metric, we have that $\|d\pi\| = 1$, because π maps points along spheres of the same radii, so the conformal factor in switching to the hyperbolic metric is the ratio of the imaginary parts of $\pi(p)$ and p, which we can compute as

$$\frac{\operatorname{Im}(\pi(p))}{\operatorname{Im}(p)} = \frac{1}{\operatorname{sech}(r)} = \cosh(r)$$

as desired.

□

Applying this lemma and utilizing the triangle inequality gives the following.

$$\begin{aligned} C^{-1}d_{\mathbb{H}^n}(x, y) - C' &\leq d_{\mathbb{H}^n}(F(x), F(y)) \\ &\leq d_{\mathbb{H}^n}(F(x), \pi_\gamma(x)) + d_{\mathbb{H}^n}(\pi_\gamma(x), \pi_\gamma(y)) + d_{\mathbb{H}^n}(\pi_\gamma(y), F(y)) \\ &\leq 2t_0 + \frac{d_{\mathbb{H}^n}(F(x), F(y))}{\cosh t_0} \\ &\leq 2t_0 + \frac{C}{C^2 + 1} d_{\mathbb{H}^n}(x, y). \end{aligned} \tag{3.6}$$

This gives us a bound λ on the length of the component $[x, y]$, as given by

$$d_{\mathbb{H}^n}(x, y) \leq \left(C^{-1} - \frac{C}{C^2 + 1}\right)^{-1} (C' + 2t_0) = \lambda. \tag{3.7}$$

Set $r = t_0 + C\lambda + 1$, which is independent of γ, and depends only on the constants C and C' which were given by the pseudo-isometry criteria of F, as desired. For any x along the path $[a,b]$, either $F(x) \in B_{t_0}(F([a,b]))$ or it is not. Notably, $B_{t_0}(F([a,b])) \subset B_r(F([a,b]))$.

For $F(x) \notin B_{t_0}(F([a,b]))$, let $[p,q]$ be a geodesic segment such that

$$x \in [a,b] \cap F^{-1}(B_{t_0}([F(a),F(b)])).$$

A bound on $F(x)$ from $[F(a),F(b)]$ is given to be below r as follows:

$$\begin{aligned} d_{\mathbb{H}^n}(F(x),[F(a),F(b)]) &\le d_{\mathbb{H}^n}(F(x),F(p)) + d_{\mathbb{H}^n}(F(p),[F(a),F(b)]) \\ &\le C d_{\mathbb{H}^n}(x,p) + t_0 \\ &\le C d_{\mathbb{H}^n}(p,q) + t_0 \\ &\le C\lambda + t_0 \\ &\le r \end{aligned} \tag{3.8}$$

so, in either case, the subset

$$F([a,b]) \subset B_r([F(a),F(b)]), \tag{3.9}$$

which is the image of a geodesic that is contained in a small ball around the geodesic between the images of endpoints.

We now show that F is a proper map, meaning the preimage of any compact set is compact. Consider some sequence q_i in γ converging to one of the end points, i.e., going to infinity in one direction. The sequence $F(q_i)$ diverges as well in $\mathbb{H}^n$. However, it has a well-defined limit point on the sphere $S^{n-1}_\infty = \overline{\mathbb{H}^n}$. Suppose otherwise, let $F(q_i')$ and $F(q_i'')$ be two subsequences that converge to different points in $\overline{\mathbb{H}^n}$. There exist indices $i,j \gg 0$ sufficiently large such that $q_i' \in [q_0'', q_j'']$ and $F(q_i') \notin B_r([F(q_0''), F(q_j'')])$ violating the previous conclusion 3.9. This is to say that these paths go to infinity, so since they stay within a bounded neighborhood of the endpoints, they cannot diverge to different points at infinity.

Let p_i and q_i be sequences converging to the opposite ends of γ as points on the sphere at infinity. Let p_∞ and q_∞ be the distinct points on $\overline{\mathbb{H}^n}$ to which $F(p_i) \to p_\infty$ and $F(q_i) \to q_\infty$ converge. Let α be the geodesic with p_∞ and q_∞ as its endpoints. Let $\gamma_i = [F(p_i), F(q_i)]$ be the geodesic segment interpolating between $F(p_i)$ and $F(q_i)$. Since $p_i \to p_\infty$ and $q_i \to q_\infty$, on some compact set $K \Subset \mathbb{H}^n$, the distance $\lim_{i\to\infty} \sup_{x\in\gamma_i\cap K} d(x,\alpha) = 0$. Take some point $x \in \gamma$ and let $K = \overline{B_{r+1}(x)}$. For $i \gg 0$ sufficiently large, the following bound holds:

$$\begin{aligned} d_{\mathbb{H}^n}(F(x),\alpha) &\le \inf_{y\in\gamma_i\cap K} (d_{\mathbb{H}^n}(F(x),y) + d_{\mathbb{H}^n}(y,\alpha)) \\ &\le \inf_{y\in\gamma_i\cap K} d_{\mathbb{H}^n}(F(x),y) + \sup_{y\in\gamma_i\cap K} d_{\mathbb{H}^n}(y,\alpha) \\ &\le r + \sup_{y\in\gamma_i\cap K} d_{\mathbb{H}^n}(y,\alpha). \end{aligned} \tag{3.10}$$

Taking the limit as $i \to \infty$, this shows that $d_{\mathbb{H}^n}(F(x),\alpha) \le r$. If we replace r with $r+1$, then γ' in the proposition statement can be realized as α and is unique by its definition given by its endpoints at infinity. $\square$

3.3 Quasi-conformal maps

We will define and state some properties of a well-behaved class of functions called *quasi-conformal* maps. Conformal maps have the property that they preserve angles, and quasi-conformal maps only distort angles in a bounded manner. As detailed in Chapter 2 in Theorem 2.3.1, the conformal maps on the boundary S^{n-1}_∞ are in bijection with isometries on the interior $\mathbb{H}^n$. Therefore, we want to associate quantities to pseudo-isometries on the interior. These will be the quasi-conformal maps on the boundary. Notably, this correspondence will not be a bijection, as we will see that pseudo-isometries are *a priori* only quasi-conformal at the boundary, but due to the ergodicity of the geodesic flow, will actually be conformal. Therefore, a pseudo-isometry and an actual isometry can both correspond to the same conformal map on the boundary, which will provide the instructions to perturb the pseudo-isometry to the unique isometry.

Definition 3.3.1 (Quasi-conformal)**.** Let $f : X \to X$ is a homeomorphism of a metric space to itself. It is K-quasi-conformal if for all $z \in X$

$$\lim_{r\to 0} \frac{\sup_{x,-x\in B(z,r)} d(f(x), f(-x))}{\inf_{x,-x\in B(z,r)} d(f(x), f(-x))} \leq K \tag{3.11}$$

for x and $-x$ antipodal points on the ball of radius r around z. f is called quasi-conformal if it is K-quasi-conformal for some K. If $K = 1$, then it is conformal. What this measures is the maximum distortion, as measured by eccentricity, of smaller disks at any point in X. Because angles are defined locally, any stretch that preserves angles cannot have any eccentricity, so the notion of 1-quasi-conformal being conformal is indicating the angle preserving feature of conformal maps.

There are two main theorems from the theory of quasi-conformal maps which we will use. The first is a differentiability result, and the second is a regularity condition as to when a quasi-conformal map is conformal.

Theorem 3.3.2. *A quasi-conformal map $f : M \to M$ for M a manifold of dimension at least two has a derivative almost everywhere.*

This theorem follows mainly from two facts in real and complex analysis: the first is that any quasi-conformal map is absolutely continuous on lines, meaning that we have a uniform modulus of continuity on any line segment L contained in M, which will give the existence of all the partial derivatives. Sometimes this is given in the definition of a quasi-conformal mapping. The second main tool in the proof is *Egorov's theorem* which takes the existence of partial derivatives and shows that almost everywhere, they combine to give total differentiability.

Theorem 3.3.3. *A quasi-conformal map whose derivative is conformal almost everywhere is conformal.*

In the two-dimensional case, locally conformality is given by holomorphic, so we can take $\mathbb{C}$ or $\mathbb{C} \cup \{\infty\}$ which exactly aligns with the hyperbolic boundary in $\mathbb{H}^3$. Here, conformal is given equivalently by the Cauchy-Riemann equations, or ϕ is conformal if and only if $\overline{\partial}\phi = 0$. If ϕ is K-quasi-conformal, then there is an inequality given by

$$|\overline{\partial}\phi| \leq k|\partial\phi|, \qquad k = \frac{K-1}{K+1}$$

where ∂ and $\overline{\partial}$ are the derivatives with respect to z and $\overline{z}$ in the complex setting. In this case, we see that if the left-hand side vanishes almost everywhere, k can be made to be 0, so integrating this result will imply that $K = 1$ and ϕ is indeed conformal.

For further details and complete proofs of these theorems, we refer the reader to Chapter 2 in Ahlfors [Ahl06].

3.4 Ergodicity

The main result we will use in Mostow's proof of rigidity is the ergodicity of the geodesic flow on hyperbolic manifolds. Intuitively, this means that generically integrating along geodesics in M will approximate behavior of functions over all of M.

Definition 3.4.1 (Measurable action)**.** Let G be a group acting on (X, μ) a (finite) measure space. We say that G is *measure class preserving* if for all measurable sets A, $\mu(A) = 0$ if and only if $\mu(gA) = 0$. This is because two measures are said to have the same *measure class* if their measure zero sets agree.

The action of G is further said to be *measure preserving* if for all measurable sets A, the measure is preserved by all G: $\mu(A) = \mu(gA)$.

We can now define the notion of *ergodic* for such actions.

Definition 3.4.2 (Ergodic)**.** A measure class preserving action G on (X, μ) is *ergodic* if any G-invariant measurable subset $A \subset X$ has either full or zero measure; that is $\mu(A) = 0$ or $\mu(X \setminus A) = 0$.

What this says is that we can use the action of G to estimate behavior over all of X. If G acts in a measure preserving manner, then G induces a unitary action on $L^2(X, \mu)$ via composition with g^{-1}, that is, each $g \in G$ acts by

$$g : L^2(X, \mu) \ni f \mapsto f \circ g^{-1}.$$

We can now characterize an ergodic action equivalently by saying the only L^2-invariant functions are constant (almost everywhere).

Theorem 3.4.3 (Ergodic $\iff$ invariant functions are constant)**.** *Suppose that (X, μ) is a finite measure space, so $\mu(X) < \infty$. A measure preserving action G is ergodic if and only if the G-invariant functions in $L^2(X, \mu)$ are exactly those that are constant almost everywhere.*

Proof. The proof follows mainly by unpacking the definitions. First, suppose that the action is not ergodic. The obstruction is therefore some subset A with non-zero measure, $\mu(A) > 0$ and $\mu(X \setminus A) > 0$ that is G-invariant. Take the function χ_A which is a G-invariant function that is not constant almost everywhere.

For the other direction, suppose that the action of G is ergodic. Given an L^2 function f which is G-invariant, consider the set

$$A_r = \{x \in X : f(x) > r\}.$$

The measure $\mu(A_r)$ must go from $\mu(X)$ to 0 as r varies along $\mathbb{R}$. If f is constant almost everywhere, then this jump happens at the value r for which f achieves almost everywhere. Otherwise, there must exist some A_r such that $0 < \mu(A_r) < \mu(X)$. This A_r would be G-invariant and therefore produce an obstruction to the ergodicity of the action. □

The critical piece of ergodic theory used to prove the ergodicity of the geodesic flow is the *von Neumann ergodic theorem* which states that ergodic actions approximate integration on the measure space.

Theorem 3.4.4 (von Neumann Ergodic theorem)**.** *Let $G = \mathbb{R}$ be a continuous ergodic action on (X, μ) a finite measure space. Let $F \subset L^2(X, \mu)$ be the G-invariant functions, which is a closed subspace since G acts by a unitary action. Let $P : L^2(X, \mu) \to F$ be the orthogonal projection operator with respect to the L^2 inner product.*

The integral of f under the action of G approximates Pf in that

$$Pf = \lim_{T \to \infty} \frac{1}{T} \int_0^T t \cdot f \, dt \tag{3.12}$$

for t considered an element of $G = \mathbb{R}$.

Proof. First we characterize conditions for a function f to be G-invariant, or what functions exactly F composes. Consider $f \in F$ and $h \in L^2(X, \mu)$ generic. Let $\langle -, - \rangle$ be the L^2 inner product. We can take an element in the subspace $\langle f, h - t \cdot h : \forall t \in G, h \in L^2(X, \mu) \rangle$ and examine its pairing with f. It is sufficient by linearity to let this be of the form $h - t \cdot h$ and we compute this inner product as

$$\begin{aligned} \langle f, h - t \cdot h \rangle &= \langle f, h \rangle - \langle f, t \cdot h \rangle \\ &= \langle f, h \rangle - \langle (-t) \cdot f, h \rangle && G \text{ acts unitarily} \\ &= \langle f, h \rangle - \langle f, h \rangle && f \text{ is } G-\text{invariant} \\ &= 0, \end{aligned}$$

which shows that all functions in $\langle h - t \cdot h \rangle$ are orthogonal to F with $\langle \cdots \rangle$ denoting the linear span of the elements.

We show that this is the full orthogonal complement to F. Suppose that f is orthogonal to $\langle h - t \cdot h : h \in L^2(X, \mu), t \in G = \mathbb{R} \rangle$ and let $h \in L^2(X, \mu)$ be arbitrary. Computing the inner product $\langle f - t \cdot f, h \rangle$ yields

$$\begin{aligned}
\langle f - t \cdot f, h \rangle &= \langle f, h \rangle - \langle t \cdot f, h \rangle \\
&= \langle f, h \rangle - \langle f, (-t) \cdot h \rangle && G \text{ acts unitarily} \\
&= \langle f, h - (-t) \cdot h \rangle \\
&= 0
\end{aligned}$$

showing that $f \in F$.

We now show Equation 3.12 for functions $f \in F$. Here we use that $Pf = f$, so we must show that

$$f = \lim_{T \to \infty} \frac{1}{T} \int_0^T t \cdot f \, dt.$$

which holds by definition. Secondly, consider elements $f = h - r \cdot h$ for any $h \in L^2(X, \mu)$ and $r \in G$. From above, we know that $Pf = 0$, so we must show

$$0 = \lim_{T \to \infty} \frac{1}{T} \int_0^T t \cdot f \, dt.$$

Since $f = h - r \cdot h$, for $T > 2r$, we can bound the integral component as

$$\begin{aligned}
\left| \int_0^T t \cdot f \, dt \right| &= \left| \int_0^T t \cdot (h - r \cdot h) \, dt \right| \\
&= \left| \int_0^T t \cdot h \, dt - \int_0^T t \cdot (r \cdot h) \, dt \right| \\
&= \left| \int_0^T t \cdot h \, dt - \int_0^r t \cdot (r \cdot h) \, dt - \int_r^T t \cdot (r \cdot h) \, dt \right| \\
&= \left| \int_0^T t \cdot h \, dt - \int_0^r t \cdot (r \cdot h) \, dt - \int_0^{T-r} t \cdot h \, dt \right| \\
&= \left| \int_{T-r}^T t \cdot h \, dt - \int_0^r t \cdot (r \cdot h) \, dt \right| \\
&= \left| \int_{T-r}^T t \cdot h \, dt - \int_0^r t \cdot h \, dt \right| \\
&\leq \left| \int_{T-r}^T t \cdot h \, dt \right| + \left| \int_0^r t \cdot h \, dt \right| \\
&\leq 2t \|h\|_{L^2}.
\end{aligned}$$

Therefore, when we normalize by multiplying by $\frac{1}{T}$, we get

$$\lim_{T\to\infty} \frac{1}{T}\int_0^T t \cdot f\, dt \le \frac{2r\|h\|}{N} \to 0,$$

completing this piece of the proof.

Extending by linearity has now shown this result on a dense set in $L^2(X,\mu)$. This is because F and $\langle h - t\cdot h\rangle$ are maximally orthogonal subspaces. Therefore, any $f \in L^2(X,\mu)$ can be arranged so that $\|f - f_0\|_{L^2} < \varepsilon/2$ and the function f_0 satisfies Equation 3.12. We can now compare Pf with the von Neumann Equation 3.12 to compute

$$\begin{aligned}
\limsup_{T\to\infty}\left\|Pf - \frac{1}{T}\int_0^T t\cdot f\, dt\right\| &\le \limsup_{T\to\infty}\left\|P(f-f_0) - \frac{1}{T}\int_0^T t\cdot (f-f_0)\, dt\right\| \\
&\quad + \limsup_{T\to\infty}\left\|Pf_0 - \frac{1}{T}\int_0^T t\cdot f_0\, dt\right\| \\
&= \limsup_{T\to\infty}\left\|P(f-f_0) - \frac{1}{T}\int_0^T t\cdot (f-f_0)\, dt\right\| \\
&\le 2\|f-f_0\| \\
&\le \varepsilon,
\end{aligned}$$

where we used that f_0 satisfies Equation 3.12 to get from line two to line three, and that P has norm at most 1 as well as $\frac{1}{T}\int_0^t t\cdot f\, dt \le \|f\|$ to get from line three to line four. $\square$

An important result is that we can achieve the same result by flowing backwards in time, i.e., the result holds equivalently to replacing the above Equation 3.12 with

$$Pf = \lim_{T\to\infty}\frac{1}{T}\int_{-T}^0 t\cdot f\, dt$$

and these forward and backwards limits will agree.

Using our discussion on the Riemannian geometry of geodesics, we know that geodesics can be given a short time existence and uniqueness, and for complete manifolds, this extends to infinite time. We use this to motivate flowing by geodesics. In this case, the input is a point on the manifold and a direction. This is parameterized by T_1M which is the unit length tangent bundle over M. Notably, this is a sphere bundle $S^{n-1} \hookrightarrow T_1M \to M$, which is a fiber bundle with fibers S^{n-1} for M a manifold of dimension n. It is compact if and only if M is compact. The unit tangent bundle inherits a Riemannian metric from (M,g) via parallel transport. To give a metric on T_1M is to give a metric on the tangent space at each point that is smoothly varying. We can represent a tangent vector at a point p in a manifold as (the equivalence class of) the derivative of a smooth curve at p. Let $\alpha, \beta : [-\varepsilon,\varepsilon] \to T_1M$ be smooth curves. These trace out paths

$$\alpha(t) = (x(t), v(t)), \quad v(t) \in T_{x(t)}M, \qquad \beta(t) = (y(t), w(t)), \quad w(t) \in T_{y(t)}M$$

that can be split into the component along M and a unit tangent direction at that point. We now use the Riemannian structure on M to compare their derivatives at $t = 0$. Define $\tilde{v}(t)$ to be the parallel transport of $v(t)$ to the point $x(0)$ along x, and similarly, let $\tilde{w}(t)$ be the parallel transport of $w(t)$ to the point $y(0)$ along y. We now can give an inner product of $\dot{\alpha}(0)$ with $\dot{\beta}(0)$ as

$$\langle \dot{\alpha}(0), \dot{\beta}(0) \rangle = \left\langle \frac{d}{dt}|_{t=0}\tilde{v}(t), \frac{d}{dt}|_{t=0}\tilde{w}(t) \right\rangle + \langle \dot{x}(0), \dot{y}(0) \rangle$$

using the Riemannian metric on M on the right side of the equation. This gives a measure on T_1M to which we can define ergodicity of the geodesic flow, defined below.

Definition 3.4.5 (Geodesic flow)**.** Given a complete compact manifold M, we define the geodesic flow

$$\mathbb{R} \times T_1M \to T_1M$$

for $(t, (x, v)) \in \mathbb{R} \times T_1M$ by taking the unique geodesic $\gamma : \mathbb{R} \to M$ such that $\gamma(0) = x$ and $\dot{\gamma}(0) = v$ and the geodesic flow maps this to $(\gamma(t), \dot{\gamma}(t)) \in T_1M$.

Our main theorem is that on a complete, finite volume hyperbolic manifold, the geodesic flow is unique.

Theorem 3.4.6 (Geodesic flow is ergodic)**.** *For M a complete, finite volume hyperbolic manifold, the geodesic flow $\mathbb{R} \times T_1M \to T_1M$ is ergodic.*

Proof. This argument is due to Hopf and is a widely used technique to study ergodic theory on various spaces. Consider T_1M to be the unit tangent bundle of M, which is an S^{n-1} bundle over M. Geodesics, normalized to unit speed, are parameterized by points in T_1M by the exponential map. We want to examine the subspace $L^2(T_1M)$ consisting of the G-invariant functions. Using the von Neumann Theorem 3.4.4 above, if we show that these functions are exactly the constant (almost everywhere) functions, then we know the flow is ergodic.

Let g_t be the geodesic flow map. First assume that $f \in C_c(T_1M)$ is a compactly supported continuous function on T_1M. This will be sufficient since such functions are dense in $L^2(X, \mu)$. The ergodic theorem tells us that

$$f_+(v) = \lim_{T\to\infty} \int_0^T f(g_t(v))\, dt$$

exists for almost all $v \in T_1M$. Similarly, we can reverse and flow backwards to get that

$$f_+(v) = \lim_{T\to\infty} \int_0^T f(g_t(v))\, dt$$

also exists for almost all $v \in T_1M$. Furthermore, these converge to the subspace of G-invariant functions under the L^2 projection operator F. Furthermore, for almost all $v \in T_1M$, we know that $f_+(v) = f_-(v) = F(v)$.

Now we utilize the geometry of hyperbolic space. Notably, if we are given two distinct geodesics that in one direction converge to the same point at infinity in S^{n-1}_∞, they diverge in the other direction. Furthermore, for any two points in the interior of $\mathbb{H}^n$, we can find geodesics through each one of them that converge to a given point on the boundary sphere at infinity.

We now pull back the geodesic flow to $T_1\mathbb{H}^n$ via the universal cover map. This can be thought of as taking a Dirichlet domain and unfolding M in $\mathbb{H}^n$, and the geodesics unfold to complete geodesics in $\mathbb{H}^n$, so they are parameterized by pairs of distinct points in S^{n-1}_∞. Let v, w be two points in T_1M such that when lifted to $T_1\mathbb{H}^n$, the geodesic flow $g_t(v)$ and $g_t(w)$ converge to the same boundary point at S^{n-1}_∞. That is, $f_+(v) = f_+(w)$ when considered on M. This means that $F(v)$ must be equal to $F(w)$, telling us that $f_-(v) = f_-(w)$. Fix v a constant. We can apply this to all w that converge in positive time to the same point at infinity as v does, which tells us that F must be constant along the $n-1$-sphere horocycle foliation of T_1M. Each one of the geodesics given by such a w converges in the negative time to a distinct point on the boundary, and by completeness, we can find these points to be dense in the boundary, (only dense because the equation of $F(v) = f_+(v)$ is only true almost everywhere). Therefore, we have shown that F must be constant almost everywhere on the boundary sphere S^{n-1}_∞.

The above analysis says that F is constant along the horocycle foliation of T_1M. This is a foliation orthogonal to the geodesic flow. We are actually applying Fubini's theorem to complete the proof that F is constant everywhere. This conclusion for compactly supported continuous functions finishes the proof upon the observation that $C_c(T_1M)$ is dense in $L^2(T_1M)$. $\square$

3.5 Extending the map to the boundary

Let F be the lift of f to the universal cover which is $\mathbb{H}^n$. Because F maps geodesics to quasi-geodesics, it induces an association of geodesics and therefore can be used to define a map between the spheres at infinity. The key insight is that a point on the sphere at infinity is an equivalence class of directed geodesics that are asymptotic or parallel. Since the above induces a unique association of geodesics, this will preserve this property of parallel and asymptotic in the universal cover, so there is a correspondence of points on the boundary. It must still be shown that F is continuous when considered as a map from $D^n \to D^n$. This is clear in the interior, so it must be shown to be continuous on the boundary.

Let x_∞ be a point on the boundary and let γ be a geodesic with an endpoint at x_∞. A local base at $F(x_\infty)$ is given as the half planes on the side of $F(x_\infty)$ of geodesics perpendicular to γ', the associated geodesic to γ. Let Q be some element of this local base. Let $x_i \to x_\infty$. At some point for $i \gg 0$ sufficiently large, all $F(x_i) \in Q$. Therefore, there exists some $x_0 \in \gamma$ such that for all $x \in [x_0, x_\infty]$, the geodesic segment from x_0 to x_∞ along γ, there exists a constant $c > 0$ from Lemma 3.6.1 such that the ball around $\pi_{\gamma'}F(x)$ is contained in Q. Let H be the hyperplane perpendicular to γ going through x_0, and Q' be the connected component of $\overline{\mathbb{H}^n} \backslash H$ containing x_∞. Let H_1 be a parallel hyperplane to H going through x. It must be that $F(H_1) \subset Q$ and so must be $F(Q') \subset Q$, thereby showing the continuity. Furthermore, to show injectivity, let $x_1 \neq x_2$ be two boundary points. Let γ be the geodesic with x_1 and x_2 as its endpoints. $F(x_1)$ and $F(x_2)$ are the endpoints of γ' and are therefore distinct. The Jordan Schoenflies theorem now says that the entire map on the sphere at infinity is a homeomorphism since it is injective and continuous. This gives our desired map $F : S_\infty^{n-1} \to S_\infty^{n-1}$, which is a homeomorphism.

3.6 Mostow's proof

We can now finish the proof of Mostow rigidity using the established background above.

Lemma 3.6.1. *Let P be a hyperplane in $\mathbb{H}^n$ and $F : \mathbb{H}^n \to \mathbb{H}^n$ be a pseudo-isometry. Let γ be a geodesic perpendicular to P and γ' be the associated geodesic given by F. There exists some $c > 0$ such that $\pi_{\gamma'}(F(P))$ lies in the region to $B_c(\gamma')$, the points of distance at most c from γ'.*

Proof. Label the point $x = P \cap \gamma$, and let y be a different point on the plane P. Let Γ be the directed geodesic starting at x and going through y. The geodesic Γ terminates at a point on the boundary y_∞ on the y-side of γ. Let Γ_1, Γ_2 be the maximal geodesics with endpoints at y_∞ and the two ends of γ. Let x_1 and x_2 be the points on Γ_1 and Γ_2 closest to x, see Figure 3.1. Because Γ and γ are perpendicular, $d(x, x_1) = d(x, x_2) = k$ is some constant independent of construction. In fact, a Möbius transformation can be applied to make the picture exactly like Figure 3.1 since they are triply-transitive. After applying the Möbius transformation, we can let x be the origin, y a pure imaginary number, and assume that γ has an endpoint to one of the sides of the horizontal geodesic.

For F, the pseudo-isometry in question, it associates to any geodesic γ the geodesic γ' as discussed. The claim is that if γ_1 and γ_2 are asymptotic, then so are γ_1' and γ_2'. This can be seen because F associates this geodesic by defining id based on its endpoints, so if γ_1' and γ_2' share an endpoint, by definition of this association, so will their images under the association map given by F. Let $z_0 = \pi_{\gamma'}(F(x))$, so from triangle inequality, the distance to Γ_1' and Γ_2' can be bounded by

$$d_{\mathbb{H}^n}(z_0, \Gamma_i') \leq d_{\mathbb{H}^n}(z_0, F(x)) + d_{\mathbb{H}^n}(F(x), F(x_i)) + d_{\mathbb{H}^n}(F(x_i), \Gamma_i') \tag{3.13}$$

and from the diagram and above argument, $d_{\mathbb{H}^n}(z_0, \Gamma_i') \leq 2r + Ck = d$, for C the constant defining F as a pseudo-isometry and the same r as in Proposition 3.2.4. Notably, d is only a function of C and C', as r is a function of C, and C' and k was independent.

The final bound of $d_{\mathbb{H}^n}(\pi_{\gamma'} F(y), z_0) \leq c$ can now be computed as

$$\begin{aligned} d_{\mathbb{H}^n}(\pi_{\gamma'} F(y), z_0) &\leq d_{\mathbb{H}^n}(\pi_{\gamma'}(F(y)), \pi_{\gamma'}(\Gamma')) + d_{\mathbb{H}^n}(\pi_{\gamma'}(\Gamma'), z_0) \\ &\leq d_{\mathbb{H}^n}(F(y), \Gamma') + d_{\mathbb{H}^n}(\pi_{\gamma'}(\Gamma'), z_0) \\ &\leq r + d_{\mathbb{H}^n}(\pi_{\gamma'}(\Gamma'), z_0) \\ &\leq r + d \end{aligned} \tag{3.14}$$

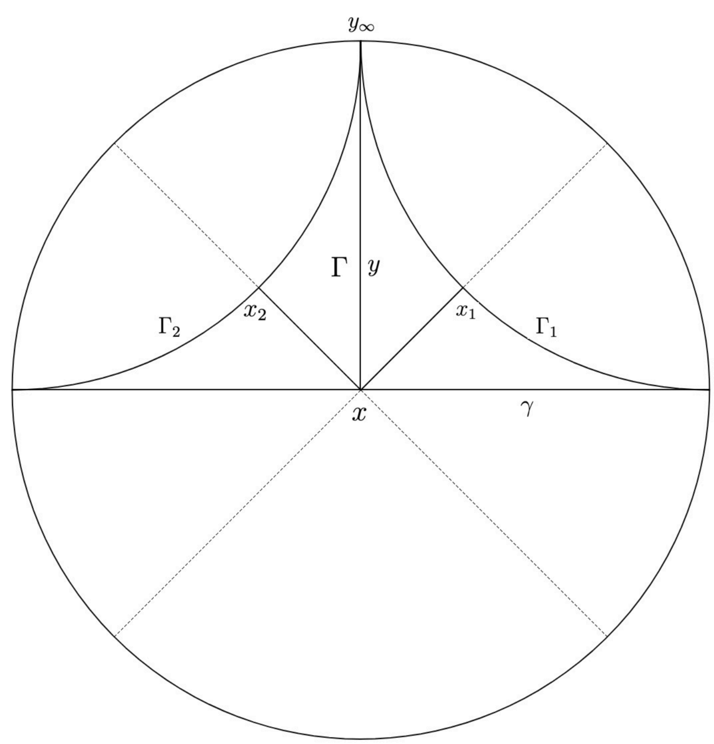

Figure 3.1: The x is on geodesic γ and Γ is a half geodesic along the plane P emanating from x in the direction of y. Γ_1 and Γ_2 are the two unique geodesics parallel to γ and Γ. x_1 and x_2 are the points on Γ_1 and Γ_2 closest to x.

completing the proof by setting $c = 2(d + r)$ since it was the diameter. $\square$

The proof is almost complete, and the rest will follow from further regularity of the map F. Firstly, it will be shown that F is quasi-conformal. This means it has a derivative almost everywhere. Where the derivative exists, it maps a sphere around 0 to an ellipsoid, so there are values $\lambda_1, \ldots, \lambda_{n-1}$ representing the eccentricities of the derivative map. Let e be the map that is the largest ratio of these eccentricities. The fundamental group of M_1 acts ergodically on $\mathbb{H}^n$, so the level set of e has full measure, and thus e is constant almost everywhere, and it turns out it will be 1 almost everywhere. A quasi-conformal map with a derivative that is conformal is actually conformal as well. This means that the now conformal map on the sphere at infinity extends to a unique isometry of $\mathbb{H}^n$, which is the map that sends $\pi_1(M_1) \to \pi_1(M_2)$ as desired.

In Chapter 2, we showed that isometries in the interior of $\mathbb{H}^n$ corresponded with conformal maps on the boundary sphere at infinity. We now establish a similar correspondence by replacing isometries with pseudo-isometries and conformal maps with quasi-conformal maps.

Proposition 3.6.2 (Pseudo-isometry on sphere at infinity is QC)**.** *If F is a pseudo-isometry from the compactified $\overline{\mathbb{H}^n}$ to itself and is a homeomorphism on the boundary, then it is quasi-conformal on the boundary sphere at infinity.*

Proof. Let x and $F(x)$ be the origin in $\mathbb{H}^n$ and $\gamma = \exp(it)$ the vertical geodesic from 0 to infinity. Let H be any hyperplane perpendicular to γ. H would look like a hemisphere centered at 0. By the normalization, assume that γ', the geodesic associated to γ, is also the vertical line from 0 to ∞. By the previous result in Lemma 3.6.1, there is a constant $c > 0$ and two hyperplanes H_1 and H_2 such that distance along γ' between any hyperplanes between H_1 and H_2 is uniformly bounded by c. This distance is seen to be the vertical distance $\log r$ for r the ratio of the radii of the S^{n-2} spheres at infinity of H_1 and H_2. The image of H lies between H_1 and H_2, so as does its boundary along S^{n-1}_∞. This means that r uniformly bounds the difference between points along S^{n-2}_H. Therefore, r is uniformly bounded and F is quasi-conformal as desired. $\square$

This is where the proof fails for hyperbolic manifolds in two dimensions. The boundary is S^1 in which quasi-conformality says almost nothing. The theorem we recall from the previous Section 3.3 states that for a manifold of dimension at least two, quasi-conformal maps have a derivative almost everywhere. This result tells us that the map F restricted to the boundary S^{n-1}_∞ has a derivative almost everywhere. Let x be a differentiable point. This means that $dF(x)$ takes a small sphere to an ellipsoid with axis lengths $\lambda_1 \leq \cdots \leq \lambda_{n-1}$. The ratio $\lambda_j \prod \lambda_i^{-1}$ is a conformal invariant. Notably, assume that this product is 1, and let $e = \lambda_{n-1}/\lambda_1$ be the maximum eccentricity of this ellipsoid. The supremum of e is the essential supremum K of the quasi-conformal mapping F.

Theorem 3.6.3. *Let M be a hyperbolic manifold of dimension at least three. $\pi_1(M)$ acts ergodically on S^{n-1}_∞ as well as on $S^{n-1}_\infty \times S^{n-1}_\infty$.*

This action comes from the universal cover map via deck transformations. In fact, it acts ergodically on $S^{n-1}_\infty \times S^{n-1}_\infty$ which is identified with the space of oriented geodesics. Ergodic means any invariant subset has either 0 or full measure.

The above theorem is simply a rephrasing of the ergodicity of the geodesic flow (Theorem 3.4.6) proven above, once we recognize that $S^{n-1}_\infty \times S^{n-1}_\infty$ is the space of geodesics. From this result, we can examine the behavior of the eccentricity map defined above. Notably, we realize that the ergodicity of the geodesics flow tells us that it is trivial.

Corollary 3.6.4. *The map e of eccentricity is constant almost everywhere.*

Proof. A level set of e is measurable. Therefore, one such level set must be of full measure. □

Let K be the constant that e is equal to almost everywhere. If $K = 1$, then the derivative of F is conformal. It will be a contradiction to have K not equal to 1. Otherwise this would create an invariant measurable set in S^{n-1}_∞, violating Theorem 3.6.3.

Finally, Theorem 3.3.3 from Section 3.3 states that this map is actually conformal because its derivative is conformal almost everywhere. Therefore, we deduce that F is a conformal map between $S^{n-1}_\infty \to S^{n-1}_\infty$ and therefore extends to a unique isometry $F : \mathbb{H}^n \to \mathbb{H}^n$ finishing the proof of rigidity in the compact setting.

3.7 Corollaries

We can now use the result of Mostow to prove some of the results claimed in the introduction, Chapter 1.

Proposition 3.7.1. *For a homotopy equivalence $f : M \to N$ between two hyperbolic manifolds of the same dimension, the isometry $\tilde{f}$ homotopic to f is unique.*

Proof. Assume that $f, g : M \to N$ are isometries and homotopic. The homotopy can be given by some $F : [0,1] \times M \to N$, and we can lift to $\tilde{F} : [0,1] \times \mathbb{H}^n \to \mathbb{H}^n$ to the universal covers. The lifts $\tilde{f}$ and $\tilde{g}$ of f and g to the universal covers are given by $\tilde{F}(0,-)$ and $\tilde{F}(1,-)$. By compactness of M and N, we can compare these two maps and get a uniform bound

$$d_{\mathbb{H}^n}(\tilde{f}(x), \tilde{g}(x)) < c$$

meaning that both $\tilde{f}$ and $\tilde{g}$ extend to the same map on the boundary sphere at infinity. Two isometries that induce the same map on the boundary are the same, a result proven in Chapter 2 Section 2.3. □

We denote $[M, M]$ to be the classes of maps from $M \to M$ up to homotopy. The subset of $[M, M]^*$ are invertible, up to homotopy. Altering any map homotopically can be realized as conjugation by an element of the fundamental group, so for an Eilenberg-Maclane space $K(G, 1)$, this group is identified with the inner automorphisms of G. The above proposition states that these are the isometries of M.

Corollary 3.7.2. *Let $M = \mathbb{H}^n/\Gamma$ be a compact hyperbolic manifold for $n \geq 3$. Then the outer automorphism group of Γ is finite.*

Proof. From the previous Proposition 3.7.1, we know that $[M, M]^* = \mathrm{Isom}(M)$. Because $M = K(\Gamma, 1)$, its isometry group is identified with the outer automorphisms of Γ.

We conclude by showing the isometry group of M is finite. Examining this map on a Dirichlet domain, it is clear that such an isometry must have a fixed point. The derivative of the map at this fixed point determines it by rigidity. The isometry group of M is compact in the C^0-topology because M is compact. Consider two elements $\phi, \psi \in \mathrm{Isom}(M)$. If the isometry group were infinite, then we could choose ϕ and ψ such that for all $x \in M$, we can bound the distance $d_{\mathbb{H}^n}(\phi(x), \psi(x))$ by the injectivity radius of M, which is finite by compactness of M. This tells us that ϕ and ψ are homotopic, so from Proposition 3.7.1, they are equal. □

We note that for dimension $n = 2$, the finiteness of the isometry group is a well-known result that the automorphism group of a Riemann surface of genus g is at most $84(g-1)$.

4

Gromov's proof of rigidity

4.1 Gromov's proof overview

The second proof of Mostow rigidity is due to Gromov, and historically takes place already knowing the result. Knowing that hyperbolic invariants descend to topological invariants, Gromov's method can be realized as first capturing the geometric notion of volume in a purely topological manner. Gromov defined a numerical quantity which is eponymously called the *Gromov norm*, which captures the complexity of the fundamental class of a manifold, a feature well-studied in topology related to integration along the manifold. The name is a slight misnomer, as this is only a semi-norm, meaning it can vanish on many non-trivial spaces including spheres. The ingenuity of his definition comes in the context of hyperbolic manifolds when this quantity exactly aligns with the hyperbolic volume, up to a normalization constant only dependent on the dimension.

We start with the same first step as in Mostow's proof, by extending the map $\tilde{f}$ to an injective map on the boundary sphere at infinity $S^{n-1}_\infty \to S^{n-1}_\infty$ continuously. The second step is to examine the volumes of simplices and ideal polyhedra, mainly providing a volume bound on such shapes, which are fundamental hyperbolic building blocks as motivated above in the construction of hyperbolic structures on knot complements. This proof reveals that encoded within the algebraic structure of the Kleinian group that defines M is the geometry of the maximal volume simplices; the algebra describes the geometry.

4.2 Gromov norm

Definition 4.2.1. Let $\alpha \in H^k(X)$ be a k-homology class. The norm $\|\alpha\|$ is defined

$$\|\alpha\| = \inf_{[\sum a_i c_i]=[\alpha]} \left\{ \sum |a_i| : \alpha = \sum a_i c_i,\ c_i \in C^k(X) \right\} \tag{4.1}$$

as the infimum over all singular chains representing α in homology of the sum of the absolute value of the coefficients.

Definition 4.2.2 (Gromov Norm)**.** Let M be a closed manifold. It has a distinguished cohomology class $[M] \in H^n(M)$ the fundamental class that generates the top homology group. The Gromov norm of M is the homology norm 4.2.1 of $[M]$ its fundamental class, $\|M\| = \|[M]\|$. The name *norm* is deceptive as for many non-contractible spaces, this quantity will vanish, but in the hyperbolic setting, this will not occur.

The remarkable discovery of Gromov is that this purely topological invariant is actually realized via geometry, in that this Gromov norm is (up to a proportionality constant) the hyperbolic volume. Firstly, there are a few inequalities of the Gromov norm that are clear by definition:

- Scalar multiplication: $\|\lambda\alpha\| \leq |\lambda|\|\alpha\|$,
- Subadditivity: $\|\alpha + \beta\| \leq \|\alpha\| + \|\beta\|$,
- Pushforward: $\|f_*\alpha\| \leq \|\alpha\|$.

All these properties come from the fact that any singular chain representing the right-hand side also represents the left, (on the third after composing with f), so the infimum on the left can be only less if there is a better representation. Notably, if $f : M \to N$ is a degree n map between spaces of the same dimension, then $\|M\| \geq |\deg(f)|\|N\|$. This gives easy computations of the Gromov norm for many manifolds. Consider $f : S^n \to S^n$ to be a degree k map; this shows that $\|S^n\| \geq k\|S^n\|$, so it must vanish.

4.3 Simplex volumes

Let $\mathcal{S}_n$ be the set of all n-dimensional hyperbolic simplices. Any simplex that is not ideal is properly contained in an ideal simplices, so notably, it does not have maximal volume. We can produce this by extending its sides. To do this, take a 1-dimensional ridge that is incomplete and complete it. Now, connect all other vertices connected to the original vertex to the new ideal vertex. This new simplex properly contains the previous one, so it has larger volume. The nature of hyperbolic geometry states that the volume of any such simplex is uniformly bounded. It is in fact true that the simplex that obtains the maximum volume is the ideal regular simplex. Ideal means each of its n-vertices lies on the sphere at infinity S^{n-1}_∞. Regular means it is maximally symmetric, like a Euclidean tetrahedron with each face an equilateral triangle, for example. Concretely, we can express the regularity of a simplex as every permutation of its vertices comes from an isometry of $\mathbb{H}^n$ that leaves the interior of the simplex invariant. All of its edges, ridges of dimension 1, have the same length. If it is not an ideal simplex, this length is finite. This constant representing the volume of an ideal regular n-simplex will be labeled v_n, the limit of the volumes of regular finite simplices of increasing edge length. The fact that this is the unique simplex obtaining the maximum volume was not known to Gromov when he proved Mostow rigidity initially, so his proof was only available for dimension $n = 3$ where it was known.

As computed in Chapter 2 Section 2.8, in dimension 2, a triangle area is determined by its angle defect $K = \pi - \alpha - \beta - \gamma$, and an ideal triangle has area π as all its angles are 0. Since the Möbius transformations are triply-transitive, all ideal triangles are congruent. For a 3-simplex, this is no longer the case. Consider $\mathbb{H}^3$ with boundary $\mathbb{C} \cup \infty$, the Riemann sphere. The isometry group consists of the Möbius transformations $\mathrm{PSL}(2, \mathbb{C})$, as it is determined by its action on the sphere at infinity. Therefore, the four vertices can be set to $0, 1, \infty, z$. The choice of z determines the congruence class of the ideal tetrahedron. It is determined by three dihedral angles that sum up to π, which is a 2-dimensional space as expected by the possibilities of the choice z. By the claim that the maximum is a regular ideal simplex, this can be given as $z = e^{i\pi/3}$, for example. In fact, in general, an ideal n-simplex can be made with a point at ∞ and n-vertices in S^{n-1}_∞ that form a Euclidean regular $(n-1)$-simplex.

Theorem 4.3.1 (Uniform bound on volume of an n-simplex)**.** *There is a uniform bound on the volume v_n of an ideal n-simplex. In fact, $v_n \leq \dfrac{\pi}{(n-1)!}$.*

Proof. By induction, the inequality claimed reduces to the claim that $v_n \leq \frac{v_{n-1}}{n-1}$. For the base case $n = 2$, the maximal volume is the ideal triangle of area π as previously computed in Chapter 2 Section 2.8. Consider an n-simplex in $\mathbb{H}^n$ with a single vertex at ∞ and the other n-vertices forming an $(n-1)$-simplex on $\mathbb{R}^{n-1} \times \{0\}$, the boundary. We can label this base τ. The volume of this can be computed exactly as in Equation 2.26 from Chapter 2 Section 2.8 over this higher dimensional base, τ. We label the coordinates of $\mathbb{H}^{n+1}$ as $(x, t) \in \mathbb{R}^n_x \times \mathbb{R}_t$ with $t > 0$ for concise notation, and the hyperbolic metric is $\frac{1}{t^2}\delta_{ij}$. The n points are all equidistant from a single point z, and the hemisphere around z lying in $\mathbb{H}^n$ forms the bottom side of the n-simplex. Label this $(n-1)$-simplex σ, and let the projection onto $\mathbb{R}^{n-1}$ be the region τ with coordinates $x = (x_1, \ldots, x_{n-1})$. The volume is the integral of 1 over the last variable t to the n, $\frac{1}{t^n}$ above the surface of the hemisphere, which we can label h. Without loss of generality, assume this hemisphere is centered at the origin and of radius 1 for the computation to appear later. Let dx be the volume form on $\mathbb{R}^{n-1}$.

$$v_n \leq \int_\tau \int_{h(x)}^{\infty} \frac{1}{t^n}\, dt\, dx = \frac{1}{n-1} \int_\tau \frac{dx}{h^{n-1}}. \tag{4.2}$$

The claim is that the integral that remains is less than the maximal volume element v_{n-1}. Let σ_0 be the $(n-1)$-simplex across from the vertex at ∞, and α be the coefficient of dilation based on the parameterized volume form of the $(n-1)$-simplex, via the map

$\tau \ni x \mapsto (x, h(x))$, and label this map ϕ, that is $\mathrm{Vol}(\sigma_0) = \int \alpha$. Showing $\alpha \geq \frac{1}{h^{n-1}}$ will complete the proof.

We notate $\langle -, - \rangle$ as the Euclidean inner product on $\mathbb{R}^{n-1}$. We can therefore express α as

$$\alpha(x) = \sqrt{\det(\langle d\phi_x(e_i), d\phi_x(e_j)\rangle)}$$

where the matrix whose determinant we are taking has entries $T_{ij} = \langle d\phi_x(e_i), d\phi_x(e_j)\rangle$. The expression of α gives the equation evaluated at v

$$d\phi_x(v) = \left(v, \frac{\langle x, v\rangle}{(1-|x|)^{\frac{1}{2}}}\right) \implies \langle d\phi_x(e_i), d\phi_x(e_j)\rangle_{\phi_x} = \frac{1}{1-|x|^2}\left(\delta_{ij} + \frac{\langle x, e_i\rangle\langle x, e_j\rangle}{1-|x|^2}\right) \quad (4.3)$$

and the determinant of this matrix, which is α^2, is computed using the fact that a matrix $A_{ij} = \delta_{ij} + a_i a_j$ has determinant $1 + \sum a_i^2$ for a_i constants. Therefore,

$$\alpha(x)^2 = \frac{1}{(1-|x|^2)^{n-1}}\left(1 + \frac{\sum\langle x, e_i\rangle^2}{1-|x|^2}\right) = \frac{1}{(1-|x|^2)^n} = \frac{1}{h(x)^{2n}} \quad (4.4)$$

and since $h \leq 1$, this shows that $\alpha(x) \geq \frac{1}{h(x)^{n-1}}$. $\square$

It was proven by Haagerup and Munkholm [HM81] in 1981 that the maximum volume simplex is uniquely achieved by a regular ideal n-simplex.

Theorem 4.3.2 (Haagerup and Munkholm)**.** *The maximum volume of a hyperbolic n-simplex is achieved only by an ideal regular n-simplex.*

See Appendix A for this result. This statement was the missing piece conjectured by Thurston to extend Gromov's proof of Mostow rigidity to all dimensions.

4.3.1 Gromov norm is hyperbolic volume

At this point, there are multiple strategies to translate between the singular homology definition of the Gromov norm and hyperbolic volume. One method, which Thurston uses, is to define the Gromov norm using homology of compactly supported and bounded total variation measures instead of the more digestible singular homology. This will make integration over the manifold easy. However, it would require a less natural topological definition of the Gromov norm, or delving more deeply into homology theory to show that these are equivalent. This approach can be seen in full detail in Thurston and Milnor [TM77] and Thurston [Thu79]. To keep within singular homology, we will need to define volume more algebraically. For this purpose, we appeal to the universal cover. Any simplex can be lifted to the universal cover using the path lifting property. Fixing the vertices, it can be made fully geodesic by replacing the simplex with the convex hull of its vertices. This does not change the simplex up to homotopy, so it has no effect on the homology class. This will align with the standard interpolations of simplices by lower dimensional simplices. Notably, any points in the interior are naturally described as convex combinations of the vertices. This process will be called *straightening* because the crooked and potentially curved sides and interior will be homotopically replaced with geodesics sweeping out the simplex. The terms and definitions utilized here can be found in Benedetti [Ben92].

Theorem 4.3.3. *The Gromov norm is greater than or equal to the volume up to proportionality,* $\|M\| \geq \frac{\mathrm{Vol}(M)}{v_n}$.

Before the proof, there are some definitions and remarks that will be useful elsewhere along the proof of the whole theorem. Let $\sigma : \Delta^k \to \mathbb{H}^n$ be a k-simplex. Recall that Δ^k is the set of points in $\mathbb{R}^{k+1}$ that form the convex hull of the standard basis e_i. Therefore, Δ^k has a natural structure as the convex hull of its vertices. Let $V_0, \dots, V_k$ be the images of the vertices under σ. A simplex in $\mathbb{H}^n$ is called straight if its image is exactly the convex hull of its vertices and the point

$$\sigma : \Delta^k \ni (t_0, \dots, t_k) \longmapsto \sum_i t_i V_i.$$

For $\pi : \mathbb{H}^n \to M$ the universal cover, a simplex in M is straight if it is realized as the projection of a straight simplex. A singular chain is straight if it is a combination of straight simplices. Notably, every chain up to homotopy can be represented by a straight chain. To see this, take $f : \Delta^k \to M$ and lift it to $\tilde{f} : \Delta^k \to M$. Every point $x \in \Delta^k$ is a convex combination of its vertices, $x = a_i v_i$ for $\sum a_i = 1$ and $a_i \geq 0$. Take the homotopy that translates the point $\tilde{f}(x)$ to the point $\sum a_i \tilde{f}(v_i)$ along the geodesic connecting them at constant speed. A simplex is degenerate if it is contained in a positive codimension hyperbolic space.

Proof of Theorem 4.3.3. This can be reduced to examining straight cycles, as any singular chain can be homotoped to be straight. Let $\sum a_i \sigma_i$. Notably, this homotopy does not change the Gromov norm of the singular chain as all the coefficients remain the same. To relate the algebraic structure to the geometric one, an algebraic definition of volume can be formed using straight simplices. Let $\varphi = \pi \circ \sigma : \Delta^n \to M$ be a straight and non-degenerate simplex. The derivative at an interior point either preserves or reverses orientation. Define the algebraic volume of φ to be the volume of σ if orientation is preserved or the negative volume if the orientation is reversed. Simply stated,

$$\mathrm{algvol}(\varphi) = \int_{\varphi(\Delta^n)} \alpha(x) dv(x) \tag{4.5}$$

for $dv(x)$ the volume form on M, and $\alpha(x)$ is $\alpha^+ - \alpha^-$ for $\alpha^+(x) = |\{t \in \mathrm{Int}(\Delta^n) : \varphi(t) = x, d\varphi_t > 0\}|$ and likewise for $\alpha^- = |\{t \in \mathrm{Int}(\Delta^n) : \varphi(t) = x, d\varphi_t < 0\}|$. This captures the algebraic multiplicity at each point. Extending linearly gives a definition of algebraic volume to all straight chains. The algebraic volume of a simplex is the hyperbolic volume of some hyperbolic simplex, so this algebraic notion of volume is also bounded by v_n for any simplex.

Let $[M]$ be represented by a straight cycle $z = \sum a_i \varphi_i$. Define the subset N as the union of all the boundary components $N = \bigcup_i \varphi_i(\partial \Delta^k)$. With $\alpha_i(x)$ as above for each simplex

φ_i, define $\Phi_z(x) = \sum a_i \alpha_i(x)$. The claim is that $\Phi_z(x) = 1$ for all $x \in M \setminus N$. Let $\mathcal{T}$ be a straight triangulation of M; first choose any triangulation, and then straighten it using the process above. This gives a canonical representation z_0 of $[M]$ that is formed from straight simplices from this triangulation. For $\Phi_{z_0}(x)$, this representation clearly demonstrated that $\Phi_z(x)$ is identically 1 on the interior of each simplex in the triangulation. This is because generic points do not lie in the $(n-1)$-skeleton of the triangulation given by $\mathcal{T}$, so Φ_{z_0} is locally constant on the interiors of the simplices, and equal to 1 since each point not in the $(n-1)$-skeleton is covered by a unique n-simplex in $\mathcal{T}$. This reduces to showing that $\Phi_{z-z_0} = 0$ on $M \setminus N$. The inclusion map $i : (M, \emptyset) \to (M, M \setminus \{x\})$ of pairs in relative homology induces an isomorphism in top homology as seen via excision and gives the identification:

$$i_* : H_n(M) \to H_n(M, M \setminus \{x\}), \qquad i_*([\omega]) = \Phi_\omega(x) \tag{4.6}$$

after fixing a generator 1 of the relative homology group $H_n(M, M \setminus \{x\})$ for $\omega = \sum b_i \omega_j$ a straight representation and x not in $\bigcup_i \omega_j(\partial \Delta^n)$. Therefore, $\Phi_{z-z_0}(x) = 0$ as desired. This map is now referred to as Φ, as its choice of representative is generically defined. Since N has positive codimension, the integral of Φ is well-defined and can be integrated to compute the volume since N does not contribute to this computation: $\mathrm{Vol}(M) = \int_M \Phi(x) dv(x)$.

Integrating the α_i terms gives the algebraic volume

$$\int_M \alpha_i(x) dv(x) = \int_{\varphi(\Delta^n)} \alpha_i(x) dv(x) = \mathrm{algvol}(\sigma_i) \tag{4.7}$$

so the hyperbolic volume can be computed

$$\mathrm{Vol}(M) = \sum a_i \mathrm{algvol}(\sigma_i) = \mathrm{algvol}\left(\sum a_i \sigma_i\right), \tag{4.8}$$

and since this was a straight simplex, the algebraic volume and hyperbolic volume in $\mathbb{H}^n$ agree, the proof is complete from

$$\mathrm{Vol}(M) = \left|\sum a_i \mathrm{algvol}(\sigma_i)\right| \leq \sum |a_i| |\mathrm{algvol}(\sigma_i)| \leq v_n \sum |a_i|. \tag{4.9}$$

□

The next step is to show the inequality in Theorem 4.3.3 is actually an equality, directly creating a link between geometry and topology.

Theorem 4.3.4. *The Gromov norm is bounded by the volume of a closed hyperbolic manifold:* $\|M\| \leq \frac{\mathrm{Vol}(M)}{v_n}$.

Corollary 4.3.5. $\|M\| = \frac{\mathrm{Vol}(M)}{v_n}$ *follows from Theorems 4.3.3 and 4.3.4.*

Example 4.3.6. The simplest case of Theorem 4.3.4 is in dimension two where it states that the volume of a closed hyperbolic surface, or a genus $g \geq 2$ surface, must be less than $4\pi(g-1)$. This is an immediate consequence of the Gauss-Bonnet theorem, which states that the volume is the Euler characteristic $2g-2$ times 2π, and since $v_2 = \pi$. To bound the Gromov norm, a genus g surface can be constructed by a $4g$-sided polygon which triangulates into $4g-2$ triangles by taking a vertex and connecting it to all vertices not already adjacent it. Each triangle is mapped onto M via a map σ_i, so $[M]$ is represented by $\sum \sigma_i$ and therefore $\|M\| \le 4g-2$.

Let M be a closed hyperbolic manifold of dimension two and consider $\Gamma = \pi_1(M)$. Since M is compact and closed, it has infinite fundamental group. Notably, it contains at least $2g$ copies of $\mathbb{Z}$. Take some copy of $\mathbb{Z}$ that is a subgroup of Γ. This subgroup gives a covering space $\pi_d : \tilde{M} \to M$ that is a d-sheeted cover corresponding to the subgroup $d\mathbb{Z} \subset \mathbb{Z}$. The Riemann-Hurwitz formula now can be used to compute $\chi(\tilde{M}) = d\chi(M)$. The genus can then be computed as $1-\chi$ to realize $g(\tilde{M}) = 1 - d(1-g)$. From above, $\|M\| \le 4g-2$, so

$$\|\tilde{M}\| \le 4(1 + d(g-1)) - 2 \le 4d(g-1) + 2$$

so dividing both sides by d yields $\|\tilde{M}\| \le 4d(g-1)-2$. Since $\tilde{M} \to M$ is a d-sheeting cover, $\|\tilde{M}\| = d\|M\|$, and this gives the bound $\|M\| \le 4(g-1) - \frac{2}{d}$. Since the Gromov norm is the infimum, this gives the desired bound as $d \to \infty$ that $\|M\| \le 4(g-1)$, completing the proof in dimension two.

To prove Theorem 4.3.4, it is necessary only now to show $\|M\| \le \frac{\mathrm{Vol}(M)}{v_n}$ given inequality in the other direction from Theorem 4.3.3.

Lemma 4.3.7. *To show the reverse inequality, $\|M\| \le \frac{\mathrm{Vol}(M)}{v_n}$, it is sufficient to prove that there exists a straight cycle $\sum a_i\sigma_i$ representing $[M]$ such that for each i,*

$$\mathrm{sgn}(a_i)\mathrm{algvol}(\sigma_i) \ge v_n - \varepsilon$$

for all $\varepsilon > 0$.

Proof. Given a straight cycle representing the fundamental class $[M] = \sum a_i\sigma_i$, we will show that the norm being arbitrarily close to v_n is equivalent to the inequality $\|M\| \le \frac{\mathrm{Vol}(M)}{v_n}$.

$$\mathrm{sgn}(a_i)\mathrm{algvol}(\sigma_i) \ge v_n - \epsilon, \quad \forall i. \tag{4.10}$$

By definition,

$$\|M\| \le \sum |a_i|. \tag{4.11}$$

Repeating the argument for the proof of Theorem 4.3.3 proves that the hyperbolic volume $\mathrm{Vol}(M) = \sum a_i \mathrm{algvol}(\sigma_i)$. Applying Inequality 4.11 gives

$$\mathrm{Vol}(M) = \sum |a_i|\mathrm{sgn}(a_i)\mathrm{algvol}(\sigma_i) \ge \sum |a_i|(v_n - \varepsilon) \ge \|M\|(v_n - \varepsilon) \tag{4.12}$$

and since the Gromov norm is the infimum, this shows the equality when combined with Theorem 4.3.3. □

We recall the results shown in Chapter 2 proven in dimensions two and three (Theorem 2.3.2) and more generally true in all dimensions, that the isometry group of $\mathbb{H}^n$ is unimodular, that is, its Haar measure is both left- and right-invariant.

Proposition 4.3.8. *The isometry group* $\mathrm{Isom}(\mathbb{H}^n)$ *is unimodular. It is naturally identified with* $\mathrm{SO}(n,1)$ *the determinant 1 matrices preserving a quadratic form of signature* $(n,1)$.

Given a Borel subset $A \subset \mathrm{Isom}(\mathbb{H}^n)$, the measure can be defined as $\mu_x(A) = \mathrm{Vol}(A(x))$ which is left-invariant by definition and right-invariant by Proposition 4.3.8. It is independent of x as well by transitivity of the isometry group.

There are a few classes of polyhedra that will be necessary. Define $\mathcal{S}(R)$ to be all fully oriented regular simplices with side length R,

$$\mathcal{S}(R) = \{(V_0, \ldots, V_n) \in (\mathbb{H}^n)^{n+1} : d_{\mathbb{H}^n}(V_i, V_j) = \delta_{ij} R\}. \tag{4.13}$$

Fully oriented means having a choice of ordering of the vertices. A polyhedron is determined uniquely by its vertices; a polyhedral simplex is the convex hull of its vertices. The distances between the vertices are invariants of the simplex up to isometries. It turns out that any regular simplex is determined (up to isometry) by the distance of any edge (recall an edge is a dimension 1 ridge). Furthermore, a simplex in $\mathbb{H}^n$ is regular if and only if $d_{\mathbb{H}^n}(V_i, V_j) = R$ is fixed for any pair of distinct vertices V_i and V_j.

Proposition 4.3.9. *A compact* n*-simplex defined by* $V_0, \ldots, V_n$ *in* $\mathbb{H}^n$ *is regular if and only if the distance between any two distinct edges is* R.

Proof. The first direction, that a regular simplex has all side lengths equal, is immediate by the definition that any permutation of the vertices arises from an isometry. Given the edge between vertices (V_i, V_j) and (V_k, V_ℓ), consider the permutation $(ik)(j\ell)$ which by assumption arises from an isometry and shows that the edge lengths are equal.

For the other direction, assume that all edge lengths are equal to R. It is sufficient to show this result only for the transposition of any two vertices, as these generate the entire symmetric group on the $n+1$ vertices. For vertices $V_i \neq V_j$, assume $0 \neq i \neq j$ and fix $V_0 = 0$ in the disk model $\mathbb{B}^n$. Define H as the linear hyperplane spanned by all vertices V_k for $k \notin \{0, i, j\}$ and the midpoint of the edge between vertices V_i and V_j. Reflection through H maintains all V_k and swaps V_i and V_j, exhibiting an isometry that realizes the transposition (ij). ☐

Elements of $\mathcal{S}(R)$ will be notated by their vertices. Let $(V_0, \ldots, V_n) \in \mathcal{S}(R)$ be a regular simplex of side length R. Any pair of such regular vertices of the same characteristic side length R are related by an isometry.

Proposition 4.3.10. *Fix a fully oriented regular geodesic simplex* $(V_0, \ldots, V_n) \in \mathcal{S}(R)$ *of side length* R*. The map*

$$\Psi : \mathrm{Isom}(\mathbb{H}^n) \to \mathcal{S}(R), \qquad \Psi(A) = (A(V_0), \ldots, A(V_n)) \tag{4.14}$$

is a bijection. That is to say that the set of fully oriented regular simplices are in correspondence with the isometry group. Notably, the isometry group acts transitively on $\mathcal{S}(R)$.

Proof. Let $A, A' \in \mathrm{Isom}(\mathbb{H}^n)$ and suppose that $\Psi(A) = \Psi(A')$. Recall that the fully oriented regular simplex $(V_0, \ldots, V_n)$ was fixed to define the map Ψ. Since Ψ maps A and A' to the same simplex V', that means each $A(V_i) = A'(V_i) = V_i'$ for all i. Therefore, $A \circ A'^{-1}$ has each V_i as a fixed point. Since there are $n+1$ such points and this map is linear, it must be the identity everywhere; therefore Ψ is injective.

To show surjectivity, assume that $V_0 = 0$ in the disk model $\mathbb{B}^n$ and that some $V' \in \mathcal{S}(R)$ has its vertices at $(V_0', V_1', \ldots, V_n')$. Since V_1' must lie on the sphere of radius R (hyperbolic radius), there exists some $A^{(1)}$ an orthogonal matrix that maps V_1 to V_1'. Therefore, assume that $V_1' = V_1$. Define $A^{(2)}$ as the orthogonal map that takes V_2 to V_1'. Repeat and $A^{(n)}$ maps V to V', showing surjectivity. □

It is clear that a simplex $V \in \mathcal{S}(R)$ has bounded volume by v_n, and as $R \to \infty$ (hyperbolic distance), the volume limits to v_n. This is seen by taking the regular Euclidean tetrahedron centered around 0 in $\mathbb{R}^n$ of Euclidean radius $r < 1$. Its vertices determine an element in $\mathcal{S}(R)$. As $r \to 1$, the hyperbolic side length $R \to \infty$ and this limits to a regular ideal simplex. The conjecture of Gromov states that such a regular ideal simplex achieves the volume v_n (see Appendix A). This result shows that $\mathrm{Vol}(V) \to v_n$ for $V \in \mathcal{S}(R)$ as $R \to \infty$. Because $\mathcal{S}(R)$ is the space of oriented regular simplices, the orientation of any element is either positive or negative. To see this, consider $V = (V_0, \ldots, V_n) \in \mathcal{S}(R)$ and then $\sigma(V_0, \ldots, V_n) : \Delta^n \ni (t_0, \ldots, t_n) \mapsto \sum t_i V_i \in \mathbb{H}^n$ is orientation preserving or reversing. Set $\mathcal{S}_+(R)$ to be the orientation preserving maps of the regular simplex of side length R, and $\mathcal{S}_-(R)$ to be the orientation reversing maps of the regular simplex of side length R. Let $\tilde{\mathcal{S}}(R)$ be the the image of the embeddings of the simplices $\mathcal{S}(R)$ in $\mathbb{H}^n$. This can be represented as $\tilde{\mathcal{S}}(R) = \mathcal{S}(R)/S_{n+1}$, the quotient of the fully oriented simplices by the symmetric group on its vertices. We are now ready to begin the proof of Theorem 4.3.4.

Using the Haar measure μ on the isometries of hyperbolic space, a measure m is defined on $\mathcal{S}(R)$ on a Borel subset $A \subset \mathcal{S}(R)$:

$$m(A) = \mu(\{\gamma \in \mathrm{Isom}(\mathbb{H}^n) : (\gamma(V_0), \ldots \gamma(V_n)) \in A\}) \tag{4.15}$$

for the fixed $V = (V_0, \ldots, V_n) \in \mathcal{S}(R)$. Since μ is left- and right-invariant, this measure is independent of choice of V.

There are some overhead definitions and constructions necessary before beginning the proof of Theorem 4.3.4. Recall that $M = \mathbb{H}^n/\Gamma$. Let $\Omega = \Gamma^{n+1}/\Gamma \cong \Gamma^n$ be defined by having Γ act on the left in each coordinate. An element of Ω can be represented by $(\mathrm{Id}, \gamma_1, \ldots, \gamma_n)$ by taking some arbitrary element $(\gamma_0, \ldots, \gamma_n) \in \Gamma^{n+1}$ and taking the representative by acting by γ_0^{-1}.

Define D as a fundamental domain for M in $\mathbb{H}^n$. We recall that we can let D be the Dirichlet domain by taking some point $m \in M$ and choosing an arbitrary preimage in $\mathbb{H}^n$ and picking all points $x \in \mathbb{H}^n$ such that $d_{\mathbb{H}^n}(m, x) \leq d_{\mathbb{H}^n}(m, \gamma(x))$ for all $\gamma \in \Gamma$, (Chapter 2 Subsection 2.6.1). Let d be the diameter of D. Since M is compact, d is finite. Fixing some interior point $V \in D \setminus \partial D$, define a simplex σ_ω for $\omega \in \Omega$ as $\pi \circ \sigma(\gamma_0(V), \ldots, \gamma_n(V))$.

The goal is to construct an explicit example of a cycle representing the fundamental class $[M]$ which has small Gromov norm. Using the definitions of σ_ω from before and choosing a representative for $\omega = (\gamma_0, \gamma_1, \dots, \gamma_n)$, define the positive and negative coefficient functions

$$a_R^{\pm}(\omega) = m((V_0', \dots, V_N') \in \mathcal{S}_{\pm}(R) : \forall j,\ V_j' \in \gamma_j(D)). \tag{4.16}$$

This is well-defined for a different representative of ω because m is secretly μ which is left-invariant with respect to the action of some $\gamma \in \Gamma$. Explicitly, recall the fixed initial $V \in \mathcal{S}(R)$ used to define $\mathcal{S}_{\pm}(R)$ with fully oriented vertices $(V_0, \dots, V_n)$:

$$\begin{aligned}
a_R^{\pm}(\omega) &= m((V_0', \dots, V_N') \in \mathcal{S}_{\pm}(R) : \forall j,\ V_j \in \gamma \circ \gamma_j(D)) \\
&= \mu(\{g \in \mathrm{Isom}^+(\mathbb{H}^n) : g(V_j) \in \gamma \circ \gamma_j(D)\}) \\
&= \mu(\{\gamma g : g \in \mathrm{Isom}^+(\mathbb{H}^n) : g(V_j) \in \gamma_j(D)\}) \\
&= \mu(\gamma\{g \in \mathrm{Isom}^+(\mathbb{H}^n) : g(V_j) \in \gamma_j(D)\}) \\
&= m(\{(V_0', \dots, V_n') \in \mathcal{S}_{\pm}(R) : V_j' \in \gamma_j(D)\})
\end{aligned}$$

utilizing the left-invariance of μ and that $\Gamma \subset \mathrm{Isom}^+(\mathbb{H}^n)$. The fundamental domain D also is such that, for any point $x_0 \in D$, the isometries $\{\delta \in \mathrm{Isom}(\mathbb{H}^n) : \delta(x_0) \in D\}$ are compact in the Lie group $\mathrm{Isom}(\mathbb{H}^n)$. and therefore has finite measure with μ, since M is compact. Therefore, $a_R^{\pm}(\omega)$ is finite.

Lemma 4.3.11. *Define $z_R = \sum(a_R^+(\omega) - a_R^-(\omega))\sigma_\omega$. This represents a non-zero multiple of $[M]$. In particular, this is a finite sum and a well-defined cycle.*

Proof. Firstly, z_R is a finite sum. Let $\omega = (\mathrm{Id}, \gamma_1, \dots, \gamma_n)$ be a chosen well-defined representative. Assume that $a_R(\omega)$ is non-zero. There is a regular simplex σ with vertices $u_0, \dots, u_n$ in $\tilde{\mathcal{S}}(R)$ such that for $u_0 \in D$, each $u_i \in \gamma_i(D)$ for all i. Let u be an interior point of the simplex, a convex combination of the u_i, so this gives the bound $d_{\mathbb{H}^n}(u, \gamma_i(u)) \leq 2d + R$ for each i where d is the diameter of D. Since Γ is discrete, only finitely many γ_i satisfy this and are possible by tesselating the region by isometric copies of d; only so many will fit this bound. Therefore, a_R^+ is 0 for a cofinite set, and therefore z_R is a finite sum.

Secondly, z_R is a cycle; it has trivial boundary as a singular chain. Consider a side τ (an $(n-1)$-dimensional face) of one of the σ_ω that is non-zero. The claim is that the coefficient of ∂z_R is 0, so it is indeed a cycle. τ is obtained as the projection of a straight $(n-1)$-simplex with vertices in the orbit of Γ acting on the vertices of u_i. That is, $\tau : \Delta^{n-1} \ni (t_0, \dots, t_{n-1}) \mapsto \pi\left(\sum t_i \gamma_i(u)\right)$ for choices γ_i and $i = 0$ to $n-1$. The coefficient of the boundary face is computed by summing over all $\gamma \in \Gamma$, the alternating sum inserting γ into each slot:

$$\sum_{j=0}^{n} (-1)^{n-j} \sum_{\gamma \in \Gamma} a_R([(\gamma_0, \dots, \gamma_{j-1}, \gamma, \gamma_j, \dots, \gamma_{n-1})]). \tag{4.17}$$

The claim is for each j, the second sum in Equation 4.17 is 0. Recall that $a_R = a_R^+ - a_R^-$, so this sum splits into these two components. Reindexing in the last slot, each component can be computed as

$$\begin{aligned}\sum_{\gamma_n\in\Gamma} a_R^{\pm}([(\gamma_0,\dots,\gamma_n)]) &= \sum_{\gamma_n} m(\{(u_0,\dots,u_n)\in\mathcal{S}_{\pm}(R) : u_i\in\gamma_i(D)\}) \\ &= m\left(\bigcup_{\gamma_n\in\Gamma}\{(u_0,\dots,u_n)\in\mathcal{S}_{\pm}(R) : u_i\in\gamma_i(D)\right) \\ &= m\{(u_0,\dots,u_n)\in\mathcal{S}_{\pm}(R) : u_i\in\gamma_i(D),\ i\le n-1,\ \exists\gamma_n\in\Gamma : u_n\in\gamma_n(D)\} \\ &= m\{(u_0,\dots,u_n)\in\mathcal{S}_{\pm}(R) : u_i\in\gamma_i(D),\ i\le n-1\}\end{aligned} \tag{4.18}$$

and from this, the sets $A_{\pm}\subset\mathcal{S}_{\pm}(R)$ can be defined as

$$A_R^{\pm} = \{(u_0,\dots,u_n)\in\mathcal{S}_{\pm}(R) : u_i\in\gamma_i(D),\ i\le n-1\} \tag{4.19}$$

and this tells that the coefficient is $m(A_R^+) - m(A_R^-)$. Let g_0 be the reflection through the hyperbolic hyperplane containing $V_0,\dots,V_{n-1}$. Recall $V=(V_0,\dots,V_n)$ was the fixed regular fully oriented n-simplex of side-length R used to define the orientation preserving and reversing simplices. Examining the positive and negative coefficient sets of $A_R^{\pm}$ can be related as

$$\{g\in\mathrm{Isom}^-(\mathbb{H}^n) : g(V_i)\in\gamma_i(D),\ i\le n-1\} = \{g\cdot g_0 : g\in\mathrm{Isom}^+(\mathbb{H}^n),\ g(V_i)\in\gamma_i(D),\ i\le n-1\} \tag{4.20}$$

so the coefficient of each boundary piece is the difference between the measures of these two sets. Since m is both left- and right-invariant, this vanishes and z_R is indeed a cycle.

It is left to be shown that z_R is a multiple of $[M]$. Suppose that $R>2d$ for d the diameter of D, a chosen fundamental domain of M in $\mathbb{H}^n$. If this is so, the $a_R^+(\omega)a_R^-(\omega)=0$ for all $\omega\in\Omega$. If $R>2d$, it means that any two adjacent vertices are in different copies of D, which tessellate the hyperbolic space $\mathbb{H}^n$. Consider some regular simplex $\sigma_0\in\tilde{\mathcal{S}}(R)$. If it has its first vertex in D, and each further vertex u_i in $\gamma_i(D)$, then any other element of $\tilde{\mathcal{S}}(R)$ must share its orientation. This is shown by choosing ω to have the identity in the first slot and noting that the regions $\gamma_i(D)$ are mutually disjoint, so there is no space for an orientation reversing map.

Recall that for any $\varepsilon>0$, if R is large enough, the algebraic volume of a regular simplex of side length R has volume at least $v_n-\varepsilon$. This was a previous lemma. It follows from the fact that the simplices achieving v_n are regular and ideal, and as $R\to\infty$, the finite length regular simplices approximate regular ideal simplices. The picture is to consider the vertices being on a Euclidean regular simplex in $\mathbb{B}^n$ on some ball of radius $r<1$ and then filling it in using hyperbolic convexity and letting $r\to 1$.

If $R>2d$, then for some ω such that $a_R(\omega)\neq 0$, we have that $a_R(\omega)\mathrm{algvol}(\sigma_\omega)>0$. Choose such an ω such that the coefficient $a_R(\omega)\neq 0$ for each $x\in M$. Define $\alpha_\omega(x)$ similar to as before, counting the positive and negative oriented points landing on x:

$$\alpha_\omega(x) = |\{t \in \mathrm{Int}(\Delta^n) : \sigma_\omega(t) = x,\ d_t\sigma_\omega > 0\}| - |\{t \in \mathrm{Int}(\Delta^n) : \sigma_\omega(t) = x,\ d_t\sigma_\omega < 0\}|. \tag{4.21}$$

The algebraic volume is computed by integrating $\alpha_\omega(x)$ over M with respect to the volume form $dv(x)$, $\mathrm{algvol}(\sigma_\omega) = \int_M \alpha_\omega(x)dv(x)$. Therefore, it must be sufficient to verify that $a_R(\omega)\alpha_\omega(x) \geq 0$.

Suppose that $a_R^+(\omega) \neq 0$ and $\alpha_\omega(x) \neq 0$. Then $\alpha_\omega(x) > 0$ since volume is positive. Take a lift $\pi : \tilde{x} \mapsto x$ and lift σ_ω using this basepoint to some $\tilde{\sigma}_\omega$ a simplex in $\mathbb{H}^n$. $\tilde{\sigma}_\omega$ is expressed as some $\sigma(\gamma_0(u), \dots, \gamma_n(u))$ for $[\gamma_0, \dots, \gamma_n]$ representing ω. Since by assumption a_R^+ is non-zero, there must be a positively oriented simplex $(u_0, \dots, u_n) \in \mathcal{S}_+(R)$ such that each $u_i \in \gamma_i(D)$. Therefore, the distance between u_i and $\gamma_i(u)$ must be within d, the diameter of D. Since $R > 2d$ by assumption, this forces $\sigma(\gamma_0(u), \dots, \gamma_n(u))$ to be positively oriented and then $\alpha_\omega(x) > 0$. The same argument shows that if a_R^- is positive, then $\alpha_\omega(x)$ is negative and the two negatives cancel as desired.

This almost completes the proof. Since $[M]$ generates the top homology, the only verification to show that z_R is a non-zero multiple of it, since it was already demonstrated to be a well-defined cycle. This follows if it is true that there is some $a_R(\omega)$ that is non-zero, as was assumed in the previous steps to show that it would be positive. Take some fully oriented simplex of side-length R $(V_0', \dots, V_n') \in \mathcal{S}(R)$. By definition, the orbits of Γ on D fill up all of $\mathbb{H}^n$. Therefore, there exist $\gamma_0, \dots, \gamma_n \in \Gamma$ such that $V_i' \in \gamma(D)$. Up to perturbing the choice of V', (or equivalently perturbing the fundamental domain D), assume that all these points are interior, that is $V_i' \in \gamma(\mathrm{Int}(D))$. From this, consider the measure m on all such simplices of this form and this therefore cannot vanish,

$$m(\{(u_0, \dots, u_n) \in \mathcal{S}(R) : u_i \in \gamma_i(D)\}) \neq 0 \tag{4.22}$$

so for any $\omega \in \Omega$, either $a_R^+(\omega) \neq 0$ or $a_R^-(\omega) \neq 0$, and therefore $a_R(\omega) \neq 0$, so this cycle does not identically vanish, completing the proof. □

We can now prove Theorem 4.3.4.

Proof of Theorem 4.3.4. Using z_R, for any $\varepsilon > 0$, take some $R > 2d$ large enough to apply the results above. It was shown that $\mathrm{sgn}(a_R(\omega))\mathrm{algvol}(\sigma_\omega) \geq v_n - \varepsilon$ whenever $a_R(\omega) \neq 0$. Choose some $k \neq 0$ such that $[z_R] = k[M]$, so $\frac{1}{k}z_R$ is homologous to $[M]$. Therefore, $\mathrm{algvol}(z_r) = k\mathrm{Vol}(M)$, so k must be positive. This is represented as a cycle by replacing each $a_R(\omega)$ with $\frac{a_R(\omega)}{k}$. This is still a straight cycle representing $[M]$ and has volume $\mathrm{sgn}\left(\frac{a_R(\omega)}{k}\right)\mathrm{algvol}(\sigma_\omega) \geq v_n - \varepsilon$. Taking $\varepsilon \to 0$ completes the reverse inequality. □

The above discussion completes a major step in the proof of Mostow rigidity, to explicitly construct the link between geometry and topology. The Gromov norm is a purely topological invariant, and this relationship to the volume shows that hyperbolic volume too is a topological invariant.

Corollary 4.3.12. *Let $f : M \to N$ be a homotopy equivalence between two hyperbolic manifolds, then* $\mathrm{Vol}(M) = \mathrm{Vol}(N)$.

Corollary 4.3.13. *Any map $f : M \to M$ must have degree 0 or ± 1.*

Proof. Any map of degree at least two forces $\|M\|$ to be 0. □

4.4 Gromov's proof

The realization of this direct bridge between geometry and topology leads to a concise proof of Mostow rigidity due to Gromov. Initially, it was only known at the time that the ideal regular simplex obtained the maximal volume v_n in dimension three, but now since that has been proven in all dimensions, his proof extends to higher dimensions as well.

Gromov's proof of Mostow rigidity. Consider the setup of a homotopy equivalence $f : M \to N$ of two closed complete hyperbolic manifolds of dimension at least three. Let F be the lift to the universal cover that extends to the sphere at infinity as detailed in the previous proof of rigidity. The main claim to finish this proof is that F takes regular ideal simplices to regular ideal simplices. The fact that allows this is not their symmetry, but rather that these are the unique simplices of maximal volume v_n.

Lemma 4.4.1. *Let $V_0, \dots, V_n$ be the vertices of a positively oriented ideal regular simplex. The points $F(V_0), \dots, F(V_n)$ form the vertices of a positively oriented ideal regular simplex.*

While this statement is true in dimension two, it is trivial. That is because all ideal 2-simplices are isometric, so this tells nothing about the structure of the map F. This lack of rigidity of ideal 2-simplices is the precise failure where the proof breaks down in dimension two. Consider A, B, C any three points on S^1_∞ the boundary of $\mathbb{H}^2$. The hyperplane from B to C is a geodesic. Consider the half of S^1_∞ not containing A. Any point along this can be considered a reflection of A through $\overline{BC}$, in that it will maintain the area. This will not be the case in higher dimensions. Let $(V_0, \dots, V_n)$ be a regular ideal n-simplex. There is a unique point V_0' such that $(V_0, \dots, V_n)$ and $(V_0', \dots, V_n)$ are both ideal regular simplices that share a side of $(V_1, \dots, V_n)$.

Proof. Suppose for the sake of contradiction that the volume of the geodesic simplex τ with vertices at $\tau = F(V_0), \dots, F(V_n)$ has volume $v_n - \varepsilon$. Assume that τ is positively oriented, perhaps by permuting two vertices if needed. Consider open sets U_i around each point V_i such that the volume of any simplex with vertices $V_0', \dots V_n'$ such that $V_i' \in U_i$ has volume at most $v_n - \varepsilon/2$. Define a chain

$$c_R = \sum_{\omega \in S} a_R(\omega)\sigma_\omega, \qquad S = \{[\gamma_0, \dots, \gamma_n] = \omega \in \Omega : \gamma_i(u) \in U_i\} \tag{4.23}$$

for u a choice of a fixed interior point of D.

Lemma 4.4.2. *There are constants $C_1, C_2 > 0$ such that for $R \gg 0$ sufficiently large, $\|z_R\| = C_1$ and $\|c_R\| \geq C_2$.*

Proof of Lemma 4.4.2. Let $V_0^R, \dots, V_n^R$ be the fixed regular simplex of side-length R used to define the orientations. These can be chosen such that $V_i^R \to w_i$ have well-defined limits to the boundary. To do this, consider one $R > 0$ and let V_i^R be symmetric around 0 in the disk model $\mathbb{B}^n$. Take all the geodesics that are radii going through 0 and V_i^R and choose each V_i^R for all R to lie on these radii.

Once R is sufficiently large such that $a_R^+(\omega)a_R^-(\omega) = 0$ for all $\omega \in \Omega$, the Gromov norm of $\|z_R\|$ can be computed as:

$$\begin{aligned}
\|z_R\| &= \sum_{\omega\in\Omega} |a_R(\omega)\| \\
&= \sum_{\omega\in\Omega} a_R^+(\omega) + a_R^-(\omega) \\
&= \sum_{\Omega\ni\omega=[\mathrm{Id},\gamma_1,\dots,\gamma_n]} m\{(u_0,\dots,u_n) \in \mathcal{S}_+(R) : u_i \in \gamma_i(D), u_0 \in D\} \\
&+ \sum_{\Omega\ni\omega=[\mathrm{Id},\gamma_1,\dots,\gamma_n]} m\{(u_0,\dots,u_n) \in \mathcal{S}_-(R) : u_i \in \gamma_i(D), u_0 \in D\} \\
&= \sum_{\Omega\ni\omega=[\mathrm{Id},\gamma_1,\dots,\gamma_n]} m\{(u_0,\dots,u_n) \in \mathcal{S}(R) : u_i \in \gamma_i(D), u_0 \in D\} \\
&= \sum_{\Omega\ni\omega=[\mathrm{Id},\gamma_1,\dots,\gamma_n]} \mu\{\delta \in \mathrm{Isom}(\mathbb{H}^n) : \delta(V_0^R) \in D, \delta(u_i) \in \gamma_i(D)\} \\
&= \mu\left(\bigcup_{\Omega\ni\omega=[\mathrm{Id},\gamma_1,\dots,\gamma_n]} \{\delta \in \mathrm{Isom}(\mathbb{H}^n) : \delta(V_0^R) \in D, \delta(u_i) \in \gamma_i(D)\}\right) \\
&= \mu\{\delta \in \mathrm{Isom}(\mathbb{H}^n) : \delta(V_0^R) \in D\} = C_1.
\end{aligned} \tag{4.24}$$

The fact that this is a constant is due to the right-invariance of μ allowed the specific choice of representative for ω to be well-defined to sum over, and independent from choice of V^R the fixed initial orientation-defining simplex. The equalities, (as opposed to inequalities), are due to R being sufficiently large.

For the second part, to bound the Gromov norm of c_R, this can be directly computed as well:

$$\|c_R\| = \sum_{\omega\in\Omega,\gamma_i(u)\in U_i} \mu\{\delta \in \mathrm{Isom}(\mathbb{H}^n) : \delta(V_i^R) \in \gamma_i(D)\} \tag{4.25}$$

and from this we can bound the norm as

$$\begin{aligned}
\|c_R\| &\geq \sum_{\omega\in\Omega,\gamma_i(D)\subset U_i} \mu\{\delta \in \mathrm{Isom}(\mathbb{H}^n) : \delta(V_i^R) \in \gamma_i(D)\} \\
&\geq \mu\left(\bigcup_{\omega\in\Omega,\gamma_i(D)\subset U_i} \mu\{\delta \in \mathrm{Isom}(\mathbb{H}^n) : \delta(V_i^R) \in \gamma_i(D)\}\right) \\
&\geq \mu\{\delta \in \mathrm{Isom}(\mathbb{H}^n) : \delta(V_i^R) \in U_i, \delta(u) \in D\}
\end{aligned} \tag{4.26}$$

where on the last line, the fixed representative of $\omega = [\mathrm{Id}, \gamma_1, \dots, \gamma_n]$ was chosen. There were multiple possible choices of representative for ω, so it might be necessary to shrink the U_i sets to U_i' around each w_i such that the closure $\overline{U_i'} \subset \mathrm{Int}(U_i)$. Define

$$\mathcal{M} = \{\delta \in \mathrm{Isom}(\mathbb{H}^n) : \delta(U_i') \subset U_i, \delta(u) \in D\}, \tag{4.27}$$

which is a neighborhood of the identity and is essentially the continuous functions from the closed $\overline{\mathbb{H}^n}$ to itself with the compact-open topology. The topology as a Lie group is not coarser than this, so it is still well-defined in the Lie group. Once $R \gg 0$ is sufficiently large, the V_i^R are close to w_i by construction and lie in U_i', and so $\mathcal{M} \subset \{\delta \in \mathrm{Isom}(\mathbb{H}^n) : \delta(V_i^R) \in U_i, \delta(u) \in D\}$, so C_2 can be set to the measure of this set $C_2 = \mu(\mathcal{M})$, completing the proof. $\square$

Using this lemma, because f is a homotopy equivalence, it must take $[M]$ to $\pm[N]$, so $\|M\| = \|N\|$, which was realized after the proof of Theorem 4.3.4 to prove Corollary 4.3.12 to show that M and N have the same volume. Recalling that $z_R = k[M]$, this states that $f_*(z_R) = \pm k[N]$. Let z_R' be the straightened representative of $f \circ z_R$. Recall that the straightening map is performed by lifting to $\mathbb{H}^n$ and replacing a chain with the chain that takes point $t \in \Delta^n$, (thought of as a convex combination of the vertices $e_0, \ldots, e_n$), to the corresponding convex combination of the vertices in $\mathbb{H}^n$ and projecting down to the hyperbolic manifold. By definition, $\mathrm{algvol}(z_R) = k\mathrm{Vol}(M)$, so then $\mathrm{algvol}(z_R') = \pm k\mathrm{Vol}(N)$. Since both M and N have the same volume, this implies that the algebraic volume of z_R and z_R' differ by ± 1.

From the previous work, for some coefficient $a_R(\omega) \neq 0$, the algebraic volume of the simplex formed by $(\gamma_0(u), \ldots, \gamma_n(u))$ for $\omega = [\gamma_0, \ldots, \gamma_n]$ is at least $v_n - \varepsilon$ and $\varepsilon \to 0$ as $R \to \infty$. Therefore, the algebraic volume of z_R is bounded below by the Gromov norm of z_R times the infimum of all such simplices:

$$|\mathrm{algvol}(z_R)| \geq \|z_R\| \inf\{\mathrm{Vol}(\sigma(\gamma_0(u), \ldots, \gamma_n(u))) : \omega = [\gamma_0, \ldots, \gamma_n] \in \Omega,\ a_R(\omega) \neq 0\}.$$

The right side limits to $\alpha_1 v_n$ as $R \to \infty$.

Let $V_F(\omega)$ be the volume of the geodesic simplex with vertices at $F(\gamma_0(u)), \ldots, F(\gamma_n(u))$. To compute the same for z_R', for F the lift to hyperbolic space as the universal cover, we compute as

$$\begin{aligned}
|\mathrm{algvol}(z_R')| &= \sum_{\omega \in \Omega} |a_R(\omega)| V_F(\omega) \\
&= \sum_{\omega : \exists i \gamma_i(u) \notin U_i} |a_R(\omega)| V_F(\omega) + \sum_{\omega : \forall i \gamma_i(u) \in U_i} |a_R(\omega)| V_F(\omega) \\
&\leq v_n \left(\sum_{\omega : \exists i \gamma_i(u) \notin U_i} |a_R(\omega)| \right) + (v_n - \varepsilon) \left(\sum_{\omega : \forall i \gamma_i(u) \in U_i} |a_R(\omega)| \right) \\
&\leq v_n(\|z_R\| - \|c_R\|) + (v_n - \varepsilon)\|c_R\| \\
&\leq C_1 v_n \left(1 - \frac{\varepsilon C_1}{v_n} C_2\right)
\end{aligned} \tag{4.28}$$

where U_i are the open sets such that if $u_i \in U_i$, the simplex spanned by the vertices u_i has volume at least $v_n - \varepsilon$, and C_1, C_2 are the constants from Lemma 4.4.2. This last inequality stated violates the fact that $\liminf |\mathrm{algvol}(z'_R)| \geq C_1 v_n$, completing the proof that F takes ideal simplices of maximal volume to ideal simplices of maximal volume. □

□

To finish Gromov's proof of Mostow rigidity, consider an isometry h such that $F \circ h|_{S^{n-1}_\infty}$ is the identity on the boundary at infinity. It would then follow that h^{-1} is homotopic to F and can therefore be used as the perturbation of f to an isometry. Let h be the isometry that takes some ideal regular simplex V to itself. Because F fixes all ideal regular simplices as shown above, take any side of V and reflect the vertex across it to another regular ideal simplex. In dimension $n \geq 3$, this is uniquely defined. Repeating this through all ideal simplices formed in this manner defines $F \circ h$ to be the identity on a dense set of the sphere at infinity and therefore everywhere.

4.5 Corollaries using the Gromov proof

Corollary 4.5.1. *Let M_1, M_2 be $\mathbb{H}^n$ quotiented by Γ_1 and Γ_2 as specific subgroups of the isometry group. For any $\phi : \Gamma_1 \to \Gamma_2$ an isomorphism, there is an isometry γ of $\mathbb{H}^n$ such that $\gamma \circ g = \phi(g) \circ \gamma$. In particular, γ induces an isometry $\tilde{f}$ between M_1 and M_2 that has $\tilde{f}_* = \phi$.*

Proof. This follows from the fact that M_1 and M_2 are Eilenberg-Maclane spaces, so there does exist a continuous map that realizes any homomorphism between the fundamental groups of M_1 and M_2. Classical Mostow rigidity finishes the proof. □

Theorem 4.5.2. *Let M and N be compact connected hyperbolic manifolds of dimension n at least three. Suppose there is a map $f : M \to N$ such that $\mathrm{Vol}(M) = |\deg(f)|\mathrm{Vol}(N)$. Then f is homotopic to a covering map that is a local isometry.*

This theorem is not an obvious corollary, as the initial step of producing a lift $\tilde{F}$ to $\mathbb{H}^n$ and extending it to the boundary is no longer continuous. It is, however, measurable, which will be sufficient to proceed. This is proven in Chapter 6 of Thurston [Thu79] for the case of dimension three, combining techniques of the geometry of simplices and some ergodic theorems. The argument could be extended to all dimensions given the results of Haagerup and Monkholm [HM81] that ideal regular simplices are the unique simplices of maximal volume, which was not known to Gromov at the time.

A

Maximal volume simplices

A.1 Regular ideal simplices have maximal volume

The maximal volume of a hyperbolic geodesic simplex is obtained uniquely by an ideal regular simplex, as conjectured by Thurston and later proven in Haagerup and Munkholm [HM81]. This theorem is clearly true in dimension two where all ideal triangles are regular. In dimension three, the computation of an ideal tetrahedron is given by the sum $\Lambda(\alpha) + \Lambda(\beta) + \Lambda(\gamma)$ for Λ the Lobachevsky function, and $\alpha + \beta + \gamma = \pi$ the dihedral angles of the tetrahedron. The Lobachevksy function is defined as

$$\Lambda(\theta) = -\int_0^\theta \log|2\sin u| du. \tag{A.1}$$

An ideal regular simplex can be expressed in $\mathbb{H}^n$ as a regular Euclidean simplex in the boundary and a point at infinity. The converse is true as well; given n points on the boundary and a point at infinity (in the upper half-space model), the n-simplex formed by them is regular if and only if the $(n-1)$-simplex in Euclidean $\mathbb{R}^{n-1}$ space is regular. Another representation is to take a regular Euclidean simplex with vertices on the unit sphere and then take the hyperbolic convex hull. This is not an if-and-only-if; an ideal simplex in $\mathbb{B}^n$ is regular if the vertices form a regular Euclidean simplex. However, as seen by reflecting one vertex through its opposite side, it is possible to get hyperbolic ideal regular simplices whose vertices do not form an ideal Euclidean simplex. In dimension two, this is clear as any three points on the unit circle work.

Proposition A.1.1. *There is a relatively tight bound on the ratio between two ideal regular simplices of adjacent dimensions:*

$$\frac{n-1}{n^2} \leq \frac{\mathrm{Vol}(V_0, \ldots, V_{n+1})}{\mathrm{Vol}(V_0, \ldots, V_n)} \leq \frac{1}{n}. \tag{A.2}$$

The upper bound was already proven in Gromov's proof of Mostow rigidity. Label $\sigma[n]$ as a regular Euclidean n-simplex with vertices on S^{n-1}_∞, as the unit sphere. The notation of $\tau[n]$ will refer to the hyperbolic simplex with these same vertices.

Define the map $p : \mathbb{B}^n \to \mathcal{K}^n$, the Poincaré (disk) model to the projective model of hyperbolic space:

$$p : \mathbb{B}^n \to \mathcal{K}^n, \qquad p : x \mapsto \frac{2}{1+|x|^2}x.$$

It is easy to work in the Klein model since the geodesics are straight lines. Note that p is not a conformal map as angles are not preserved in the Klein model. The metric can be computed in the Klein model as

$$g_{ij} = \frac{1}{1-|x|^2}\delta_{ij} + \frac{x_i x_j}{(1-|x|^2)^2},$$

and its volume form is therefore

$$dV = \frac{1}{(1-|x|^2)^{\frac{n+1}{2}}}\, dx_1 \cdots dx_n.$$

This allows us to integrate over $\sigma[n] = p(\tau[n])$, which in this model represents an ideal regular hyperbolic simplex. The volume is therefore the standard integral using the above volume form. Therefore, the volume of an ideal regular simplex is computed as

$$\mathrm{Vol}(\tau[n]) = \int_{p(\tau[n])} \frac{dx}{(1-|x|^2)^{\frac{n+1}{2}}}. \tag{A.3}$$

Lastly, recall the map $h : \mathbb{B}^n \to \mathbb{H}^n$, the isometry between the Poincaré disk model and the upper half space model

$$h : (x_1, \ldots, x_n) \mapsto \frac{1}{|(x_1, \ldots, x_{n-1}, x_n - 1)|^2}(2x_1, \ldots, 2x_{n-1}, 1-|x|^2)$$

which takes the point $(0, \ldots, 0, 1)$ to the point ∞. Using this can give another computation of the volume of the simplex $\tau[n]$ knowing the volume form on $\mathbb{H}^n$ is $x_n^{-n}\, dx_1 \ldots dx_n$. Use an isometry to let $V_0 = (0, \ldots, 0, 1)$, which maps to ∞ in $\mathbb{H}^n$, and let the other vertices $V_1, \ldots, V_n$ be the vertices of a regular Euclidean simplex in $S^{n-2} \subset \mathbb{R}^{n-1}_\infty$ the boundary of the upper half-space model. The total n-simplex can be thought of as vertically stacked $(n-1)$-simplices. Consider the map ϵ taking the model $(n-1)$-simplex formed by $V_1, \ldots, V_n$ in $\mathcal{K}^n$ to the regular Euclidean simplex in S^{n-2} in $\mathbb{H}^n$. The image $h(\tau[n]) \setminus \{\infty\}$

can be thought of as $\epsilon(\tau[n]) \times [0, \infty)$. This gives a formula for its volume, integrating over this as:

$$\mathrm{Vol}(\tau[n]) = \int_{\epsilon(\tau[n])} \int_{\sqrt{(1-\rho^2)}}^{\infty} x_n^{-n}\, dx d\rho = \frac{1}{n-1} \int_{\epsilon(\tau[n])} (1-\rho^2)^{\frac{-1-n}{2}} d\rho \tag{A.4}$$

for $\rho^2 = x_1^2 + \cdots x_{n-1}^2$ the radius function on the sphere at infinity $\mathbb{R}_\infty^{n-1}$.

Proof of Proposition A.1.1. Let $\sigma_0[n]$ be any regular Euclidean simplex, and $\tau_0[n]$ be the corresponding hyperbolic simplex with the same vertices. Consider the following three integrals whose combination directly implies the result:

(a) $\int_{\sigma_0[n]} \frac{1}{(1-|x|^2)^{\frac{n+1}{2}}}\, dx = \mathrm{Vol}(\tau_0[n])$

(b) $\int_{\sigma_0[n]} \frac{1}{(1-|x|^2)^{\frac{n}{2}}}\, dx = n\mathrm{Vol}(\tau_0[n+1])$

(c) $\int_{\sigma_0[n]} \frac{1}{(1-|x|^2)^{\frac{n-1}{2}}}\, dx = \frac{n-1}{n}\mathrm{Vol}(\tau_0[n])$

By the inequalities of the integrands, once the equalities stated are proven, it is implied that

$$\frac{n-1}{n}\mathrm{Vol}(\tau_0[n]) \leq n\mathrm{Vol}(\tau_0[n+1]) \leq \mathrm{Vol}(\tau_0[n])$$

finishing the proof.

Up to the action of an isometry, we can assume that $p(\tau_0[n])$ is a regular Euclidean simplex. The first integral in (a) is proven by the integral computation in Equation A.3 and the second in (b) is similarly the computation from Equation A.4, because the fact that $\tau_0[n+1]$ is regular implies that $\epsilon(\tau_0[n+1])$ is regular as well. To compute integral (c), it is easier to work on the boundary using Stoke's theorem. In this setting, the explicit variation is described by the divergence of a vector field:

$$\int_{\sigma_0[n]} \mathrm{div}\ V(x)\, dx = \int_{\partial\sigma_0[n]} V \cdot n\, dS \tag{A.5}$$

for dS the surface volume form on the boundary.

Applying the divergence theorem from Equation A.5 to the vector field $V = \frac{x}{(1-|x|^2)^{\frac{n-1}{2}}}$ will demonstrate the result. Computing each partial derivative yields

$$\frac{\partial V_i}{\partial x_i} = \frac{(1-|x|^2)^{\frac{n-1}{2}}) + (n-1)x_i(1-|x|^2)^{\frac{n-3}{2}} x_i)}{(1-|x|^2)^{n-1}} = (1-|x|^2)^{-\frac{n-1}{2}} + (n-1)x_i^2(1-|x|^2)^{-\frac{n+1}{2}} \tag{A.6}$$

so the divergence is calculated as the sum over i of Equation A.6 giving

$$\begin{aligned}\text{div } V &= \frac{n}{(1-|x|^2)^{\frac{n-1}{2}}} + (n-1)\frac{|x|^2}{(1-|x|^2)^{\frac{n+1}{2}}} \\ &= \frac{n}{(1-|x|^2)^{\frac{n-1}{2}}} + (n-1)\left(\frac{1}{(1-|x|^2)^{\frac{n+1}{2}}} - \frac{1-|x|^2}{(1-|x|^2)^{\frac{n+1}{2}}}\right) \\ &= \frac{1}{(1-|x|^2)^{\frac{n-1}{2}}} + (n-1)\frac{1}{(1-|x|^2)^{\frac{n+1}{2}}}.\end{aligned} \tag{A.7}$$

Labeling $\phi_n(\alpha) = \int_{\sigma_0[n]}(1-|x|^2)^{-\alpha}\,dx$, the integral of the divergence computed above is simply expressed as $\phi_n\left(\frac{n-1}{2}\right) + (n-1)\phi_n\left(\frac{n+1}{2}\right)$.

The boundary of $\sigma_0[n]$ consists of $n+1$ faces, each of which is a regular Euclidean $(n-1)$-simplex. Label the face opposite to vertex i as $\partial_i\sigma_0[n]$. Each such boundary component has its vertices in a sphere S^{n-2} the Euclidean sphere in the hyperplane containing $\partial_i\sigma_0[n]$ and let $\rho_n = \sqrt{1-n^{-2}}$ be the radius of this sphere. In this component, $(1-|x|^2) = \rho_n^2 - \rho^2$ where ρ is the distance from the center of the boundary component to the point x. Furthermore, the dot product $x \cdot n$ is computed as $\frac{1}{n}$. This easily can be seen by the fact that since n is normal to the boundary component, the component of x in the direction of n is the same for all x in this component, so setting $x = n = \frac{1}{\sqrt{n}}(1,\ldots,1,0)$ for $i = n+1$ computes this dot product at all points.

The right-hand side of Equation A.5 can be computed as

$$\frac{n+1}{n}\int_{\partial_0\sigma_0[n]}(\rho_n^2-\rho^2)^{-\frac{n-1}{2}}\,d\rho \tag{A.8}$$

since each boundary component will have the same integral by the regularity assumption. Furthermore, each boundary component is itself a regular simplex, so this integral can be re-expressed in a lower dimension as

$$\frac{n+1}{n}\int_{\sigma_0[n-1]}(\rho_n^2-\rho_n^2|x|^2)^{-\frac{n-1}{2}}\rho_n^{n-1}\,dx = \frac{n+1}{n}\int_{\sigma_0[n-1]}(1-|x|^2)^{-\frac{n-1}{2}}\,dx = \frac{n+1}{n}\phi_{n-1}\left(\frac{n-1}{2}\right) \tag{A.9}$$

since the boundary component is the rescaled simplex of one dimension lower with scaling term defined as ρ_n. Combining these results gives the inductive formula

$$\phi_n\left(\frac{n-1}{2}\right) + (n-1)\phi_n\left(\frac{n+1}{2}\right) = \frac{n+1}{2}\phi_{n-1}\left(\frac{n-1}{2}\right). \tag{A.10}$$

Using the equations of parts (a) and (b), and substituting in the volume of an ideal regular simplex, the equality in part (c) is proven by

$$\phi_n\left(\frac{n+1}{2}\right) = \text{Vol}(\tau_0[n]), \qquad \phi_{n-1}\left(\frac{n-1}{2}\right) = (n-1)\text{Vol}(\tau_0[n]) \tag{A.11}$$

which implies

$$\phi_n\left(\frac{n-1}{2}\right) = \frac{n-1}{n}\mathrm{Vol}(\tau_0[n]), \tag{A.12}$$

finishing the proof of part (c). ☐

There is one final lemma which will nearly complete the proof that only ideal regular simplices obtain the maximal volume of a simplex.

Lemma A.1.2. *Let $f : (0,1] \to \mathbb{R}$ be a concave function. Let p be the center of mass of $\sigma[n]$ an arbitrary Euclidean simplex with vertices at $V_0, \dots, V_n \in S^{n-1}$. Denote $c = |p|$ the norm of the center of mass. When the following (improper) integrals converge, they satisfy the inequality*

$$A := \frac{1}{\mathrm{Vol}(\sigma[n])}\int_{\sigma[n]} f(1-|x|^2)\,dx \leq \frac{1}{\mathrm{Vol}(\sigma_0[n])}\int_{\sigma_0[n]} f((1-c^2)(1-|x|^2))\,dx =: B. \tag{A.13}$$

When f is strictly concave, (meaning its second derivative is strictly negative), then the inequality is sharp if and only if $\sigma[n]$ is regular.

Note that when $\sigma[n]$ is regular, $(1-c^2) = 1$, so the integrands are the same up to an isometry taking the vertices of $\sigma_0[n]$ to $V_0, \dots, V_n$ the vertices of $\sigma[n]$.

Proof. Let $\Delta[n]$ be the standard simplex of points $\Delta[n] = \{(t_0, \dots, t_n) : t_i \geq 0, \sum t_i = 1\}$. Consider the map $(t_0, \dots, t_n)$ mapping to the convex combination $\sum t_i V_i$, giving a homeomorphism $\Delta[n] \to \sigma[n]$. Let μ be the Lebesgue probability measure of $\Delta[n]$, and it therefore descends to a probability measure on $\sigma[n]$. This recomputes A as an integral over the standard simplex as

$$A = \int_{\Delta[n]} f(1-|\sum t_i V_i|^2)d\mu$$

and since μ is the Euclidean measure, it is invariant under any isometries of $\mathbb{R}^{n+1}$, notably any permutation of the vertices. Therefore, $A = \int_{\Delta[n]} f(1-|\sum t_i V_i|^2)d\mu = \int_{\Delta[n]} f(1 - |\sum t_{\rho(i)} V_i|^2)d\mu$ for $\rho \in S_{n+1}$ a permutation on $(n+1)$ vertices. We can take the average over ρ iterating through all possible permutations, and using the concavity of f, this gives the inequality

$$A \leq \int_{\Delta[n]} f\left(\frac{1}{(n+1)!}\sum_{\rho \in S_{n+1}} \left(1-|\sum t_{\rho(i)} V_i|^2\right)\right) d\mu$$

To compute the integrand, the symmetry will be used to cancel out many terms or take the sum to obtain c, the center of mass.

$$|\sum_i t_i V_i|^2 = \sum_{i,j} t_{\rho(i)} r_{\rho(j)} \langle V_i, V_j\rangle = \sum_{i\neq j} t_{\rho(i)} r_{\rho(j)} \langle V_i, V_j\rangle + \sum t_i^2 \tag{A.14}$$

and averaging over all of S_{n+1} gives

$$\frac{1}{(n+1)!}\sum_{\rho\in S_{n+1}} t_{\rho(i)}t_{\rho(j)} = \frac{1}{n(n+1)}\sum_{k\neq\ell} t_k t_\ell = \frac{1}{n(n+1)}\left(1-\sum t_i^2\right), \qquad i\neq j \tag{A.15}$$

and lastly the sum over all inner products of V_i is

$$\sum\langle V_i, V_j\rangle = |\sum V_i|^2 - \sum |V_i|^2 + n + 1 = (n+1)^2c^2. \tag{A.16}$$

Plugging these formulae into the previous inequality gives

$$A \leq \int_{\Delta[n]} f\left(\frac{n+1}{n}(1-c^2)(1-\sum t_i^2)\right) d\mu \tag{A.17}$$

and if $\sigma[n]$ is regular, then $c = 0$ and each permutation doesn't affect any of the sums, so the above is an equality. Taking Equation A.17 and substituting in $g(x) = f((1-c^2)x)$ and setting $\sigma[n]$ to $\sigma_0[n]$ yields

$$B = \int_{\Delta[n]} f\left(\frac{n+1}{n}(1-c^2)(1-\sum t_i^2)\right) d\mu \tag{A.18}$$

showing the inequality $A \leq B$ as desired.

Suppose that $A = B$. As discussed, this clearly occurs when $\sigma[n]$ is regular. In this case, the inequality A.17 must be an equality as well. When f is strictly concave, this can only occur when $|\sum t_{\rho(i)}V_i|^2 = |\sum t_iV_i|^2$ for each i and ρ. Plugging in particular values can compute these terms. Letting $t_0 = t_1 = \frac{1}{2}$ and all the rest 0, shows that $|V_1+V_2| = |V_i+V_j|$ for all $i\neq j$. The parallelogram law states that $|V_i-V_j|^2 = 4-|V_i+V_j|^2$ and gives similarly that $|V_1 - V_2| = |V_i - V_j|$ for all $i \neq j$. This is exactly the statement that $\sigma[n]$ is regular, finishing the proof. □

Theorem A.1.3. *A hyperbolic n-simplex has maximal volume v_n if and only if it is ideal and regular.*

Proof. This proof proceeds by induction. It is clearly true for dimensions two and three. In dimension two, all ideal triangles are regular. In dimension three, the volume of ideal simplices is computed with the Lobachevsky function Λ, which is maximized when it is regular.

Define $K_n = \dfrac{n\mathrm{Vol}(\tau_0[n+1])}{\mathrm{Vol}(\tau[n])}$. For $\tau[n+1]$ an arbitrary ideal simplex, consider the function $f(t) = t^{-\frac{n}{2}} - K_n t^{-\frac{n+1}{2}}$, which will be a function to plug into the results above. For $n \geq 3$, this function is indeed strictly concave on $(0, 1]$ if and only if $K_n \geq \frac{n(n+2)}{(n+1)(n+3)}$. This now satisfies the assumptions to apply Lemma A.1.2 to f and the Euclidean simplex

$\sigma[n] = \epsilon(\tau[n+1])$. Using Equations A.3 and A.4 for $n+1$, let $\tau[n] = p^{-1}(\sigma[n])$ to compute

$$\begin{aligned} n\mathrm{Vol}(\tau[n+1]) - K_n \mathrm{Vol}(\tau[n]) &\le \int_{\sigma_0[n]} f((1-c^2)(1-|x|^2))\,dx \\ &= (1-c^2)^{-\frac{n}{2}} n\mathrm{Vol}(\tau_0[n+1]) - K_n (1-c^2)^{-\frac{n+1}{2}} \mathrm{Vol}(\tau_0[n]) \\ &\le (1-c^2)^{-\frac{n}{2}} (n\mathrm{Vol}(\tau_0[n+1]) - K_n \mathrm{Vol}(\tau_0[n])) \\ &= 0. \end{aligned} \tag{A.19}$$

The inductive hypothesis at this point states that $\mathrm{Vol}(\tau[n]) \le \mathrm{Vol}(\tau_0[n])$ since the volume is maximized at a regular simplex. The previous computation A.19 therefore gives the inequality

$$n\mathrm{Vol}(\tau[n+1]) \le K_n \mathrm{Vol}(\tau_0[n]) = n\mathrm{Vol}(\tau_0[n+1]) \tag{A.20}$$

showing that the volume of $\tau_0[n+1]$ is maximal since $\tau[n+1]$ was arbitrary. If the inequality in A.20 is sharp, then it is also an equality in Equation A.19. Lemma A.1.2 then says that $\epsilon(\tau[n+1])$ is Euclidean regular, and therefore $\tau[n+1]$ is hyperbolically regular, finishing the proof. □

B
Algebraic topology

B.1 Eilenberg-Maclane spaces

We prove an important theorem from algebraic topology that we used in the proof of Mostow rigidity.

Proposition B.1.1. *Let $\rho : \pi \to \pi'$ be any group homomorphism. There exists a unique map up to homotopy $f \in [K(\pi, n), K(\pi', n)]$ such that $f_* : \pi_n(K(\pi, n)) \to \pi_n(K(\pi', n))$ is ρ.*

Corollary B.1.2. *For M and N hyperbolic manifolds with the same fundamental group π, any automorphism ρ of π can be realized by a map $f : M \to N$ such that $\rho = f_*$ is the induced map on $\pi_1(M) \to \pi_1(N)$.*

Proof of Proposition B.1.1. If $n = 0$, then this is just a collection of discrete points indexed by the groups and this is trivial. Firstly, the Eilenberg-Maclane spaces do exist. The construction involved requires the *Moore space* $M(\pi, n)$, which has reduced homology trivial except in degree n where it is π, and the taking $P^n M(\pi, n)$, the n^{th}-Postnikov section which kills off all higher homotopy groups. The Moore space is the mapping cone (or cofibration) of the map given by

$$\bigvee_{j \in J} S^n \xrightarrow{f} \bigvee_{i \in I} S^n$$

where f is the map on spheres given by a presentation of π as

$$\bigoplus_{j \in J} \mathbb{Z} e'_j \xrightarrow{f} \bigoplus_{i \in I} \mathbb{Z} e_i \to \pi$$

via a free resolution. By the Hurewicz theorem, $M(\pi, n)$ has trivial homotopy groups until degree n, where it will agree with the homology group and be π. The n^{th}-Postnikov section kills off all higher homotopy groups and realizes the Eilenberg-Maclane space.

Let $X = K(\pi, n)$ and $Y = K(\pi', n)$. For every $f \in [X, Y]$ we define a homomorphism $\rho : \pi \to \pi'$ by mapping $f : X \to Y$ to the unique homomorphism $\phi : \pi \to \pi'$ fitting into the diagram:

$$\begin{tikzcd}
\pi \arrow[r, dotted, "\phi"] \arrow[d, "i"] & \pi' \arrow[d, "j"] \\
\pi_n(X) \arrow[r, "f_*"] & \pi_n(Y)
\end{tikzcd}.$$

The construction of the Moore space and Postnikov sections gives rise to $K(\pi, n)$ as a CW complex. Let $X^{(m)}$ be the m-skeleton. The attaching maps to create the $(m+1)$-skeleton are given by the pushout square

$$\begin{tikzcd}
\bigvee S^m \arrow[r] \arrow[d] & \bigvee D^{m+1} \arrow[d] \\
X^{(m)} \arrow[r] & X^{(m+1)}
\end{tikzcd}. \tag{B.1}$$

Since the disk D^{m+1} can be considered as the cone over its boundary S^m, the wedge has the same property $\bigvee D^{m+1} \cong C\left(\bigvee S^m\right)$. This expresses $X^{(m+1)}$ as $X^{(m)} \cup C\left(\bigvee S^m\right)$ glued along the attaching map. Consider now the map

$$Y^{X^{(m)}} \to Y^{\vee S^m} \tag{B.2}$$

which will have homotopy fiber given by $Y^{X^{(m+1)}}$ from above.

Taking the long exact sequence of homotopy groups of the above fibration yields the exact sequence

$$\left[\bigvee S^{m+1}, Y\right] \to [X^{(m_1)}, Y] \to [X^{(m)}, Y] \to \left[\bigvee S^m, Y\right]. \tag{B.3}$$

Interpreting exactness when these are the π_0 is as follows: Since the first term in sequence B.3 is the fundamental group $\pi_1\left(Y^{\vee S^m}\right)$, it acts on maps $X^{(m+1)} \to Y$. Exactness means that such maps are in the same orbit of $\pi_1\left(Y^{\vee S^m}\right)$ if and only if they map to the same element in $[X^{(m)}, Y]$.

Once $m > n$, this sequence degenerates since X and Y have no homotopy groups anymore, so we deduce that

$$[X, Y] \to [X^{(n+1)}, Y] = [M(\pi, n), Y] \tag{B.4}$$

is a bijection, and for $m = n$ this is exactly

$$0 \to [M(\pi, n), Y] \to \operatorname{Hom}\left(\bigoplus_{j \in J} \mathbb{Z}e'_j, \pi'\right) \to \operatorname{Hom}\left(\bigoplus_{i \in I} \mathbb{Z}e_i, \pi'\right). \tag{B.5}$$

The result follows by applying the Whitehead theorem stating that a weak equivalence in the category of CW complexes is an actual homotopy equivalence. □

References

[Ahl06] Lars Ahlfors. *Lectures on Quasiconformal Mappings.* University lecture series (Providence, R.I.) ; 38. American Mathematical Society, Providence, R.I., 2nd ed. edition, 2006.

[Ano69] Dmitrij Viktorovič Anosov. *Geodesic flows on closed Riemann manifolds with negative curvature*, volume 90 of *Proceedings of the Steklov Institute of mathematics.* American Mathematical Society, Providence - R.I, 1969.

[Ben92] Riccardo Benedetti. *Lectures on Hyperbolic Geometry.* Universitext. Springer-Verlag, Berlin, Heidelberg, 1st ed. 1992. edition, 1992.

[Ber53] Lipman Bers. *Theory of Pseudo-analytic Functions.* New York University. Institute for Mathematics and Mechanics, New York, 1953.

[Bon09] Francis Bonahon. *Low-Dimensional Geometry : from Euclidean Surfaces to Hyperbolic Knots.* Student mathematical library v. 49. American Mathematical Society ; Institute for Advanced Study, Providence, R.I. : Princeton, N.J., 2009.

[Bow93] Brian H. Bowditch. Geometrical finiteness for hyperbolic groups. *Journal of functional analysis*, 113(2):245–317, 1993.

[Car92] Manfredo Perdigão do Carmo. *Riemannian Geometry.* Mathematics (Boston, Mass.). Birkhäuser, Boston, 1992.

[DM69] Pierre Deligne and David Mumford. The irreducibility of the space of curves of given genus. *Publications mathématiques. Institut des hautes études scientifiques*, 36(1):75–109, 1969.

[Don11] Simon K. Donaldson. *Riemann Surfaces.* Oxford graduate texts in mathematics Riemann surfaces. Oxford University Press, Oxford, 2011.

[Gro06] Michael Gromov. Hyperbolic manifolds according to Thurston and Jørgensen. pages 40–53, 2006.

[Hat02] Allen Hatcher. *Algebraic Topology.* Cambridge University Press, Cambridge ; New York, 2002.

[HM81] Uffe Haagerup and Hans J. Munkholm. Simplices of maximal volume in hyperbolic n-space. *Acta mathematica*, 147:1–11, 1981.

[HM98] Joe Harris and Ian Morrison. *Moduli of Curves*. Graduate Texts in Mathematics, 187. Springer, 1st ed. 1998. edition, 1998.

[Hub06] John H. Hubbard. *Teichmüller Theory and Applications to Geometry, Topology, And Dynamics. Volume 1 Teichmüller Theory 1*. Matrix Press, 2006.

[Lac00] Marc Lackenby. Notes for graduate hyperbolic manifolds course, Hilary term. *University of Oxford*, 2000.

[McM92] Curtis T. McMullen. Riemann surfaces and the geometrization of 3-manifolds. *Bulletin (new series) of the American Mathematical Society*, 27(2):207–216, 1992.

[McM96] Curtis T. McMullen. *Renormalization and 3-Manifolds which Fiber over the Circle*. Annals of mathematics studies ; Number 142. Princeton University Press, Princeton, New Jersey, 1996.

[Mos67] George D. Mostow. On the rigidity of hyperbolic space forms under quasiconformal mappings. *Proceedings of the National Academy of Sciences - PNAS*, 57(2):211–215, 1967.

[Ota14] Jean-Pierre Otal. Three-dimensional manifolds, Kleinian groups and hyperbolic geometry. *Jahresbericht der Deutschen Mathematiker-Vereinigung*, 116(1):3–20, 2014.

[Rat94] John G. Ratcliffe. *Foundations of Hyperbolic Manifolds*. Graduate texts in mathematics ; 149. Springer Science+Business Media, LLC, New York, 1st ed. 1994. edition, 1994.

[Spa98] Ralf Spatzier. Lectures on spaces of nonpositive curvature (and ergodicity of geodesic flows). *Bulletin of the London Mathematical Society*, 30(3):317–335, 1998.

[Thu79] William P. Thurston. *The Geometry and Topology of Three-Manifolds*. s.n., s.l., 1979.

[Thu82] William P. Thurston. Three dimensional manifolds, Kleinian groups and hyperbolic geometry. *Bulletin (New Series) of the American Mathematical Society*, 6(3):357 – 381, 1982.

[Thu97] William P. Thurston. *Three-dimensional Geometry and Gopology*. Princeton mathematical series ; 35. Princeton University Press, Princeton, N.J., 1997.

[TM77] William P. Thurston and John Milnor. Characteristic numbers of three-manifolds. *L'Enseignement Mathématique*, 23, 1977.