CHAPTER 2: HYPERBOLIC GEOMETRY

DANNY CALEGARI

Abstract. These are notes on hyperbolic 3-manifolds, which are being transformed into Chapter 2 of a book on 3-Manifolds. These notes follow courses given at the University of Chicago in Spring 2015, Spring 2021 and Fall 2024.

Contents

1. Models of hyperbolic space

1.1. **Trigonometry.** The geometry of the sphere is best understood by embedding it in Euclidean space, so that isometries of the sphere become the restriction of *linear* isometries of the ambient space. The natural parameters and functions describing this embedding and its symmetries are transcendental, but satisfy algebraic differential equations, giving rise to many complicated identities. The study of these functions and the identities they satisfy is called *trigonometry*.

In a similar way, the geometry of hyperbolic space is best understood by embedding it in *Minkowski* space, so that (once again) isometries of hyperbolic space become the restriction of *linear* isometries of the ambient space. This makes sense in arbitrary dimension, but the essential algebraic structure is already apparent in the case of 1-dimensional spherical or hyperbolic geometry.

1.1.1. *The circle and the hyperbola.* We begin with the differential equation

$$f''(\theta) + \lambda f(\theta) = 0 \tag{1.1}$$

for some real constant λ, where f is a smooth real-valued function of a real variable θ. The equation is 2nd order and linear so the space of solutions V_λ is a real vector space of

Date: January 21, 2025.

dimension 2, and we may choose a basis of solutions $c(\theta), s(\theta)$ normalized so that if $W(\theta)$ denotes the Wronskian matrix

$$W(\theta) := \begin{pmatrix} c(\theta) & c'(\theta) \\ s(\theta) & s'(\theta) \end{pmatrix} \tag{1.2}$$

then $W(0)$ is the identity matrix.

Since the equation is autonomous, translations of the θ coordinate induce symmetries of V_λ. That is, there is an action of (the additive group) $\mathbb{R}$ on V_λ given by

$$t \cdot f(\theta) = f(\theta + t)$$

At the level of matrices, if $F(\theta)$ denotes the column vector with entries the basis vectors $c(\theta)$, $s(\theta)$ then $W(t)F(\theta) = F(\theta + t)$; i.e.

$$\begin{pmatrix} c(t) & c'(t) \\ s(t) & s'(t) \end{pmatrix} \begin{pmatrix} c(\theta) \\ s(\theta) \end{pmatrix} = \begin{pmatrix} c(\theta + t) \\ s(\theta + t) \end{pmatrix} \tag{1.3}$$

If $\lambda = 1$ we get $c(\theta) = \cos(\theta)$ and $s(\theta) = \sin(\theta)$, and the symmetry preserves the quadratic form $Q_E(xc + ys) = x^2 + y^2$ whose level curves are circles. If $\lambda = -1$ we get $c(\theta) = \cosh(\theta)$ and $s(\theta) = \sinh(\theta)$, and the symmetry preserves the quadratic form $Q_M(xc+ys) = x^2 - y^2$ whose level curves are hyperbolas. Equation 1.3 becomes the angle addition formulae for the ordinary and hyperbolic sine and cosine.

We parameterize the curve through $(1, 0)$ by $\theta \to (c(\theta), s(\theta))$. This is the parameterization by angle on the circle, and the parameterization by *hyperbolic length* on the hyperboloid.

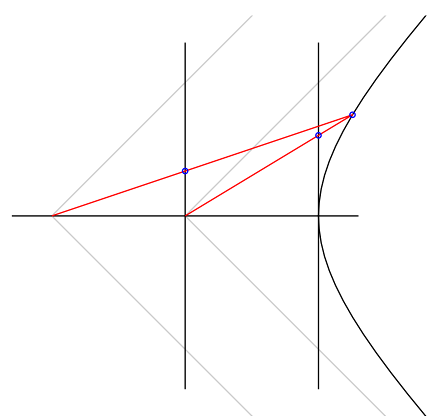

FIGURE 1. Projection to the tangent and stereographic projection to the y axis takes the point $(\cosh(\theta), \sinh(\theta))$ on the hyperboloid to the points $(1, \tanh(\theta))$ on the tangent and $(0, \tanh(\theta/2))$ on the y-axis.

1.1.2. *Projection to the tangent.* Linear projection from the origin to the tangent line at $(1, 0)$ takes the coordinate θ to the *projective* coordinate $t(\theta)$ (which we abbreviate t for simplicity). This is a degree 2 map, and we can recover $c(\theta), s(\theta)$ up to the ambiguity of sign by extracting square roots. For the circle, $t = \tan$ and for the hyperbola $t = \tanh$.

The addition law for translations on the θ-line becomes the addition law for ordinary and hyperbolic tangent:

$$\tan(\alpha+\beta) = \frac{\tan(\alpha)+\tan(\beta)}{1-\tan(\alpha)\tan(\beta)}; \quad \tanh(\alpha+\beta) = \frac{\tanh(\alpha)+\tanh(\beta)}{1+\tanh(\alpha)\tanh(\beta)} \tag{1.4}$$

1.1.3. *Stereographic projection.* Stereographic linear projection from $(-1,0)$ to the y-axis takes the coordinate θ to a coordinate $\rho(\theta) := s(\theta)/(1+c(\theta))$. This is a degree 1 map, and we can recover $c(\theta), s(\theta)$ algebraically from ρ. The addition law for translations on the θ-line expressed in terms of ρ_E for the circle and ρ_M for the hyperboloid, are

$$\rho_E(\alpha+\beta) = \frac{\rho_E(\alpha)+\rho_E(\beta)}{1-\rho_E(\alpha)\rho_E(\beta)}; \quad \rho_M(\alpha+\beta) = \frac{\rho_M(\alpha)+\rho_M(\beta)}{1+\rho_M(\alpha)\rho_M(\beta)} \tag{1.5}$$

The only solutions to these functional equations are of the form $\tan(\lambda\theta)$ and $\tanh(\lambda\theta)$ for constants λ, and in fact we see $\rho_E(\theta) = \tan(\theta/2)$ and $\rho_M(\theta) = \tanh(\theta/2)$.

1.2. **Higher dimensions.** We now consider the picture in higher dimensions, beginning with the linear models of spherical and hyperbolic geometry.

1.2.1. *Quadratic forms.* In $\mathbb{R}^{n+1}$ with coordinates $x_1, \cdots, x_n, z$ define the quadratic forms Q_E and Q_M by

$$Q_E = z^2 + \sum x_i^2 \text{ and } Q_M = -z^2 + \sum x_i^2$$

We can realize these quadratic forms as symmetric diagonal matrices, which we denote Q_E and Q_M without loss of generality. For Q one of Q_E, Q_M we let $\mathrm{O}(Q)$ denote the group of linear transformations of $\mathbb{R}^{n+1}$ preserving the form Q.

In terms of formulae, a matrix M is in $\mathrm{O}(Q)$ if $(Mv)^T Q(Mv) = v^T Qv$ for all vectors v; or equivalently, $M^T QM = Q$. Denote by $\mathrm{SO}^+(Q)$ the connected component of the identity in $\mathrm{O}(Q)$. If $Q = Q_E$ then this is just the subgroup with determinant 1. If $Q = Q_M$ this is the subgroup with determinant 1 and lower right entry > 0.

We also use the notation $\mathrm{SO}(n+1)$ and $\mathrm{SO}^+(n,1)$ for $\mathrm{SO}^+(Q)$ if we want to stress the signature and the dependence on the dimension n.

Example 1.1. If $n = 1$ then $\mathrm{SO}^+(Q)$ is 1-dimensional, and consists of Wronskian matrices $W(\theta)$ as in equation 1.2.

We let S denote the hypersurface $Q_E = 1$ and H the sheet of the hypersurface $Q_M = -1$ with $z > 0$. If we use X in either case to denote S or H then we have the following observations:

Lemma 1.2 (Homogeneous space)**.** *The group* $\mathrm{SO}^+(Q)$ *preserves* X*, and acts transitively with point stabilizers isomorphic to* $\mathrm{SO}(n,\mathbb{R})$*.*

Proof. The group $\mathrm{O}(Q)$ preserves the level sets of Q, and the connected component of the identity preserves each component of the level set; thus $\mathrm{SO}^+(Q)$ preserves X.

Denote by p the point $p = (0, \cdots, 0, 1)$. Then $p \in X$ and its stabilizer acts faithfully on T_pX which is simply $\mathbb{R}^n$ spanned by $x_1, \cdots, x_n$ with the standard Euclidean inner product. Thus the stabilizer of p is isomorphic to $\mathrm{SO}(n,\mathbb{R})$, and it remains to show that the action is transitive.

This is clear if $Q = Q_E$. So let $(x, z) \in H$ be arbitrary. By applying an element of $\mathrm{SO}(n, \mathbb{R})$ (which acts on the x factor in the usual way) we can move (x, z) to a point of the form $(0, 0, \cdots, 0, x_n, z)$ where $x_n = \sinh(\tau)$, $z = \cosh(\tau)$ for some τ. Then the matrix

$$A(-\tau) := I_{n-1} \oplus W(-\theta) = I_{n-1} \oplus \begin{pmatrix} \cosh(-\tau) & \sinh(-\tau) \\ \sinh(-\tau) & \cosh(-\tau) \end{pmatrix} \tag{1.6}$$

takes the vector $(0, 0, \cdots, 0, x_n, z)$ to p. $\square$

Denote by A_H the subgroup of $\mathrm{SO}^+(Q_M)$ consisting of matrices $A(\tau)$ as above, and by A_S the subgroup of $\mathrm{SO}(Q_E)$ consisting of matrices $I_{n-1} \oplus W(\theta)$, and denote either subgroup by A. Similarly, in either case denote by K the subgroup $\mathrm{SO}(n, \mathbb{R})$ stabilizing the point $p \in X$. Note that A_H is isomorphic to $\mathbb{R}$, whereas A_S is isomorphic to S^1. Then we have the following:

Proposition 1.3 (KAK decomposition). *Every matrix in* $\mathrm{SO}^+(Q)$ *can be written in the form* $k_1 a k_2$ *for* $k_1, k_2 \in K$ *and* $a \in A$. *The expression is unique up to* $k_1 \to k_1 k$, $k_2 \to k^{-1} k_2$ *where* k *is in the centralizer of* a *intersected with* K *(which is the upper-diagonal subgroup* $\mathrm{SO}(n-1, \mathbb{R})$ *unless* a *is trivial).*

Proof. Let $g \in \mathrm{SO}^+(Q)$ and consider $g(p)$. If $g(p) \neq p$ there is some $k_2 \in K$ which takes $g(p)$ to a vector of the form $(0, 0, \cdots, x_n, z)$, where k_2 is unique up to left multiplication by an upper-diagonal element of $\mathrm{SO}(n-1, \mathbb{R})$. $\square$

It is useful to spell out the relationship between matrix entries in $\mathrm{SO}^+(Q)$ and geometric configurations. Any time a Lie group G acts on a Riemannian manifold M by isometries, it acts freely on the Stiefel manifold $V(M)$ of orthonormal frames in M, so we can identify G with any orbit. When M is homogeneous and isotropic, each orbit map $G \to V(M)$ is a diffeomorphism. In this particular case, the diffeomorphism is extremely explicit:

Lemma 1.4 (Columns are orthonormal frames). *A matrix* M *is in* $\mathrm{SO}^+(Q)$ *if and only if the last column is a vector* v *on* X, *and the first* n *columns are an (oriented) orthonormal basis for* $T_v X$.

Proof. This is true for the identity matrix, and it is therefore true for all M because $\mathrm{SO}^+(Q)$ acts by left multiplication on itself and on X, permuting matrices and orthonormal frames. It is transitive on the set of orthonormal frames by Proposition 1.3. $\square$

1.2.2. *Distances and angles.* Since the restriction of the form Q to the tangent space TX is positive definite, it inherits the structure of a Riemannian manifold. The group $\mathrm{SO}^+(Q)$ acts on X by isometries.

Note if $v \in X$, then we can identify the tangent space $T_v X$ with the subspace of $\mathbb{R}^{n+1}$ consisting of vectors w with $w^T Q v = 0$; it is usual to denote this space by $v^\perp$. For the basepoint p, we can identify $T_p X$ with the Euclidean space spanned by the x_i. Thus for any two vectors $a, b \in T_p X$ we have

$$\cos(\angle(a, b)) = \frac{a^T Q b}{\|a\|\|b\|} \tag{1.7}$$

Since the action of $\mathrm{SO}^+(Q)$ preserves angles and inner products, this formula is valid for any two vectors $a, b \in v^\perp = T_v X$ at any $v \in X$.

Similarly, if $v, w \in X$ are any two points, there is some $g \in \mathrm{SO}^+(Q)$ and some $A(\tau)$ so that $g(v) = p$ and $g(w) = A(\tau)(p)$. Now, $A'(0) \in T_pX$ and $\|A'(0)\| = 1$ so the curve $\tau \to A(\tau)(p)$ is parameterized by arclength. The upper-diagonal subgroup $\mathrm{SO}(n-1, \mathbb{R})$ fixes precisely this curve pointwise, so it must be totally geodesic. In particular, in this example, $d(v, w) = \tau$, so that

$$c(d(v,w)) = \frac{v^T Q w}{\|v\|\|w\|}$$

where c denotes cosh or cos in the hyperbolic or spherical case, and we use the convention that $\|v\| = i$ for v on the positive sheet of $Q_M = -1$. To see this, use the fact that both sides are invariant under the action of $\mathrm{SO}^+(Q)$, and compute in the special case $v = p$, $w = A(\tau)(p)$, $d(v, w) = \tau$. In the spherical case, this formula reduces to equation 1.7. In the hyperbolic case, it is given by

$$\cosh(d(v,w)) = \frac{v^T Q w}{\|v\|\|w\|} \tag{1.8}$$

1.2.3. *Sine and cosine rule.* Three points A, B, C on X span a geodesic triangle with angles α, β, γ and lengths a, b, c (where a is the length of the edge opposite the angle α at point A and so on). Three generic points span a 3-dimensional subspace of $\mathbb{R}^{n+1}$, so without loss of generality we may take $n = 2$ throughout this section.

It is convenient to introduce the notation of the *dot product* $u \cdot v := u^T Q v$ and the *cross product*, defined by the formula $(u \times v) \cdot w = \det(uvw)$.

After an isometry, we can move the vectors A, B, C to the points

$$A = (0, 0, 1), \quad B = (s(c), 0, c(c)), \quad C = (s(b)\cos(\alpha), s(b)\sin(\alpha), c(b))$$

where s, c are $\sinh, \cosh$ or $\sin, \cos$ depending on whether we are in the hyperbolic or spherical case. By equations 1.7 and 1.8 we obtain the *cosine rule*

$$c(a) = \frac{B \cdot C}{\|B\|\|C\|} = c(b)c(c) \pm s(b)s(c)\cos(\alpha)$$

Explicitly, in spherical geometry this gives

$$\cos(a) = \cos(b)\cos(c) + \sin(b)\sin(c)\cos(\alpha) \tag{1.9}$$

and in hyperbolic geometry this gives

$$\cosh(a) = \cosh(b)\cosh(c) - \sinh(b)\sinh(c)\cos(\alpha) \tag{1.10}$$

Using the same coordinates for A, B, C we obtain the following formula for the determinant:

$$(A \times B) \cdot C = \det(ABC) = s(b)s(c)\sin(\alpha)$$

But matrices in $\mathrm{SO}^+(Q)$ have determinant 1 so this must be symmetric in cyclic permutations of A, B, C and therefore

$$s(b)s(c)\sin(\alpha) = s(c)s(a)\sin(\beta) = s(a)s(b)\sin(\gamma)$$

dividing through by $s(a)s(b)s(c)$ we obtain the *sine rule*. Explicitly in spherical geometry this gives

$$\frac{\sin(\alpha)}{\sin(a)} = \frac{\sin(\beta)}{\sin(b)} = \frac{\sin(\gamma)}{\sin(c)} \tag{1.11}$$

and in hyperbolic geometry this gives

$$\frac{\sin(\alpha)}{\sinh(a)} = \frac{\sin(\beta)}{\sinh(b)} = \frac{\sin(\gamma)}{\sinh(c)} \tag{1.12}$$

1.2.4. *Geodesics and geodesic subspaces.* The geodesic through p which is the orbit of the subgroup $A(\tau)$ is precisely the intersection of X with the 2-plane $\pi_0 := \{x_1 = x_2 = \cdots = x_{n-1} = 0\}$. This 2-plane is spanned by $p \in X$ and $a := (0, \cdots, 0, 1, 0) \in T_pX$. Since $\mathrm{SO}^+(Q)$ acts transitively on the unit tangent bundle of X, every geodesic in X is the intersection of X with a 2-plane π; the 2-planes that intersect X are precisely those on which the restriction of Q is indefinite and nondegenerate. The stabilizers of geodesics are the subgroups conjugate to $\mathrm{SO}(n-1) \times A$ which is equal to $\mathrm{SO}(n-1) \times \mathrm{SO}^+(1,1)$ or $\mathrm{SO}(n-1) \times \mathrm{SO}(2)$.

Similarly, the intersection of X with the $k+1$-plane $\{x_1 = x_2 = \cdots = x_{n-k} = 0\}$ is a totally geodesic subspace of dimension k, and all such subspaces arise this way. The stabilizers are the subgroups conjugate to $\mathrm{SO}(n-k) \times \mathrm{SO}^+(k,1)$ or $\mathrm{SO}(n-1) \times \mathrm{SO}(k+1)$.

1.2.5. *Klein projective model.* Projection from the origin to the tangent plane $z = 1$ at the point $(0, \cdots, 0, 1)$ takes H to the interior of the unit ball B in $z = 1$. The group $\mathrm{SO}^+(Q_H)$ acts faithfully by projective linear transformations. This defines the *Klein projective model* of hyperbolic space. In this model, hyperbolic straight lines and planes are the intersection of Euclidean straight lines and planes with B. The plane $z = 1$ can be compactified to real projective space $\mathbb{RP}^n$. B is thus a convex domain in $\mathbb{RP}^n$ bounded by a quadric, and the group of hyperbolic isometries is the same as the group of projective transformations of $\mathbb{RP}^n$ preserving a quadric.

For $n = 2$ a quadric in $\mathbb{RP}^2$ is the image of an $\mathbb{RP}^1$ under a degree 2 embedding (the Veronese embedding) which is stabilized by a copy of $\mathrm{PSL}(2, \mathbb{R})$ in $\mathrm{PSL}(3, \mathbb{R})$ obtained by projectivizing S^2V, the symmetric square of the standard representation V of $\mathrm{SL}(2, \mathbb{R})$. This exceptional case is discussed again in § 1.3.2.

If ℓ is a projective line over any field, a projective automorphism of ℓ preserves the *cross-ratio* of an ordered 4-tuple $x, y, z, w \in \ell$, which is the ratio

$$(x, y; z, w) := \frac{(x-z)(y-w)}{(y-z)(x-w)}$$

There are 24 ways to permute the 4 entries. If λ is the cross ratio of one permutation, the various permutations take the 6 values

$$\lambda, \quad \frac{1}{\lambda}, \quad \frac{1}{1-\lambda}, \quad 1-\lambda, \quad \frac{\lambda}{\lambda-1}, \quad \frac{\lambda-1}{\lambda}$$

and each of these six values is also sometimes called a "cross-ratio".

Suppose p, q are points in B. There is a unique maximal straight line segment ℓ in B containing p and q, and intersecting ∂B at $\ell(0)$ and $\ell(1)$. Then there is a formula for the hyperbolic distance from p to q in terms of a cross ratio:

$$d_K(p, q) = \frac{1}{2} \log \frac{(q - \ell(0))(\ell(1) - p)}{(p - \ell(0))(\ell(1) - q)} \tag{1.13}$$

In the special case that p is the origin, and q is a point in the disk at Euclidean radius r corresponding to hyperbolic distance θ, we obtain

$$\theta := d_K(p,q) = \frac{1}{2}\log\frac{1+r}{1-r} \tag{1.14}$$

To see this, observe that projective automorphisms of an interval preserve the cross-ratio, and therefore equation 1.13 reduces to equation 1.14. But we have already seen (by our analysis of the 1-dimensional case) that $r = \tanh(\theta)$, which is equivalent to equation 1.14.

1.2.6. *Poincaré unit ball model.* Stereographic projection from $(0,\cdots,0,-1)$ to the plane $z = 0$ also takes H to the interior of the unit ball B in $z = 0$. This is a *conformal* model, in the sense that it is angle-preserving. Geodesics in the Poincaré model are straight lines through the origin and arcs of round circles perpendicular to ∂B.

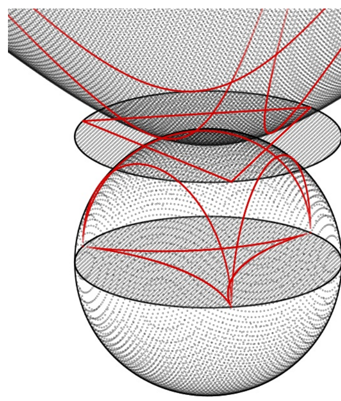

FIGURE 2. An ideal triangle in various models.

First we show that geodesics are straight lines through the origin and round circles perpendicular to the boundary. To see this, we factorize stereographic projection from H to $z = 0$ in three steps. First, we perform projection from the origin to the plane $z = 1$; the image is the Klein model, whose straight lines are Euclidean straight lines. Second, we project vertically to the unit sphere S; thus hyperbolic straight lines are taken to round circles on the sphere perpendicular to the equator. Finally, stereographic projection from the south pole to $z = 0$ is conformal and takes hyperbolic straight lines to round circles perpendicular to ∂B.

The fact that this composition of maps agrees with direct stereographic projection from the hyperboloid to $z = 0$ can be seen by a direct computation. Since all maps are symmetric with respect to $\mathrm{SO}(n,\mathbb{R})$ (the stabilizer of $(0,\cdots,0,1)$) we can restrict attention to a typical point $(0,\cdots,0,\sinh(\theta),\cosh(\theta))$ on a single radial geodesic. For brevity we only write the last two coordinates. The three projections (which we denote K, v and s) map

$$(\sinh(\theta),\cosh(\theta)) \xrightarrow{K} (\tanh(\theta),1) \xrightarrow{v} \left(\tanh(\theta),\frac{1}{\cosh(\theta)}\right) \xrightarrow{s} \left(\frac{\sinh(\theta)}{\cosh(\theta)+1},0\right)$$

Now let's show the Poincaré disk model is conformal; i.e. that the projection $\pi : H \to B$ from the hyperboloid to the unit ball takes orthonormal frames in H (in the hyperbolic metric) to perpendicular frames of equal length in B (in the Euclidean metric). The easiest way to compute hyperbolic distances between points in B in the Poincaré model is to project back to the hyperboloid by

$$(x_1, \cdots, x_n, 0) \to \left(\frac{2x_1}{1-\sum x_i^2}, \cdots, \frac{2x_n}{1-\sum x_i^2}, \frac{1+\sum x_i^2}{1-\sum x_i^2} \right) \tag{1.15}$$

and use equation 1.8. In the special case that p is the origin, and q is a point at radius r we obtain

$$\theta := d_P(p,q) = \log \frac{1+r}{1-r} \tag{1.16}$$

which recovers $r = \rho(\theta) = \tanh(\theta/2)$ as we obtained in the 1-dimensional case. After a symmetry fixing $p = 0$, we can suppose q is on the x_n axis, corresponding to the point $q' := (0, \cdots, 0, \sinh(\theta), \cosh(\theta))$ on the hyperboloid. By Lemma 1.4, one orthonormal frame at q' is given by vectors

$$v_1 := (1, 0, \cdots, 0), \quad v_2 := (0, 1, 0, \cdots, 0), \quad \cdots, \quad v_n := (0, \cdots, 0, \cosh(\theta), \sinh(\theta))$$

and we deduce that a hyperbolic circle with radius θ has perimeter $2\pi\sinh(\theta)$. By symmetry, the vector $d\pi(v_j)$ for $j < n$ is perpendicular to the (Euclidean) sphere of radius r centered at the origin, and thus (by comparing perimeters of circles) it has (Euclidean) length $r/\sinh(\theta) = \tanh(\theta/2)/\sinh(\theta)$. On the other hand, the projection $d\pi(v_n)$ is tangent to the radius, and its (Euclidean) length is $dr(\theta)/d\theta$ which, by equation 1.16 is $d\tanh(\theta/2)/d\theta = 1/(2\cosh^2(\theta/2))$. But

$$\|d\pi(v_j)\| = \frac{\tanh(\theta/2)}{\sinh(\theta)} = \frac{\sinh(\theta/2)}{\cosh(\theta/2)(2\sinh(\theta/2)\cosh(\theta/2))} = \frac{1}{2\cosh^2(\theta/2)} = \|d\pi(v_n)\|$$

In particular, the vectors of the frame $\{d\pi(v_i)\}$ are mutually perpendicular and of the same (Euclidean) length, so that π is conformal as claimed.

Differentiating with respect to r, and using the fact that the model is conformal, we can express the Riemannian length element ds_P (in the hyperbolic metric) in terms of the usual Euclidean metric ds_E on B by the formula

$$ds_P = \frac{2ds_E}{1-r^2} \tag{1.17}$$

1.2.7. *Upper half-space model.* Inversion in a tangent sphere takes the unit ball conformally to the upper half-space; in n dimensions with coordinates $x_1, \cdots, x_{n-1}, z$ the upper half-space is the open subset $z > 0$. In this model, the hyperbolic metric ds_P is related to the Euclidean metric ds_E by the formula

$$ds_P = \frac{ds_E}{z} \tag{1.18}$$

Hyperbolic straight lines in this model are round circles and straight lines perpendicular to $z = 0$.

The "planes" $z = C$ for $C > 0$ a constant are called *horospheres*. These are the horospheres centered at ∞; other horospheres in this model are round Euclidean spheres tangent to some point in $z = 0$.

In its intrinsic (Riemannian) metric, a horosphere is isometric to Euclidean space $\mathbb{E}^{n-1}$, although it is exponentially distorted in the extrinsic metric. The group $\mathbb{R}^{n-1}$ acts by translations on $z = 0$ and simultaneously on all the horospheres $z = C$ (although the translation length depends on C); we denote this subgroup of $\mathrm{SO}^+(Q)$ by N. If we choose coordinates where the axis of the subgroup A is the vertical line $x_1 = x_2 = \cdots = x_{n-1} = 0$, then this group acts as dilations centered at the origin. As before, let $K = \mathrm{SO}(n;\mathbb{R})$ denote the stabilizer of a point on the axis of A; without loss of generality we can take the point $(0, \cdots, 0, 1)$. Then we have the following:

Proposition 1.5 (KAN decomposition)**.** *Every matrix in* $\mathrm{SO}^+(Q)$ *can be written* uniquely *in the form kan for* $k \in K$*,* $a \in A$ *and* $n \in N$*.*

Proof. Let $p = (x_1, \cdots, x_{n-1}, z)$ in the upper half-space be arbitrary. We first move p to $(0, \cdots, 0, z)$ by horizontal translation by the vector $n^{-1} := (-x_1, \cdots, x_{n-1}) \in \mathbb{R}^{n-1} = N$. Then move it to $(0, \cdots, 0, 1)$ by a dilation $a^{-1} \in A$ centered at 0 which scales everything by $1/z$. The composition moves p to $(0, \cdots, 0, 1)$. Since K is the stabilizer of $(0, \cdots, 0, 1)$, we are done. □

1.3. **Dimension 2 and 3.** Some exceptional isomorphisms of Lie groups in low dimensions allow us to express the transformations in the conformal models especially simply.

1.3.1. *Unit disk and upper half-plane.* If we identify $\mathbb{R}^2$ with $\mathbb{C}$ conformally, then hyperbolic automorphisms in the unit disk and unit half-plane models become holomorphic automorphisms of the Riemann sphere.

Thinking of the Riemann sphere as the complex projective line $\mathbb{CP}^1$, the group of automorphisms is just $\mathrm{PGL}(2,\mathbb{C}) = \mathrm{PSL}(2,\mathbb{C})$, acting projectively by

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot z = \frac{az+b}{cz+d} \tag{1.19}$$

The subgroup fixing the unit circle is $\mathrm{PSU}(1,1)$, whose elements are represented (uniquely up to sign) by matrices of the form $\left(\begin{smallmatrix} \alpha & \beta \\ \bar{\beta} & \bar{\alpha} \end{smallmatrix}\right)$ with $|\alpha|^2 - |\beta|^2 = 1$. The subgroup fixing the real line is $\mathrm{PSL}(2,\mathbb{R})$, whose elements are represented (uniquely up to sign) by real matrices of the form $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ with $ad - bc = 1$. These subgroups are conjugate in $\mathrm{PSL}(2,\mathbb{C})$, and this conjugacy relates the Poincaré unit disk and upper half-plane models.

The dynamics of an isometry can be expressed in terms of its trace (which is only well-defined up to sign). Fix an isometry g, expressed as a matrix in $\mathrm{SL}(2,\mathbb{C})$ which is unique up to multiplication by -1.

1. If $|\mathrm{tr}(g)| < 2$ then g is *elliptic*. It fixes a unique point in the interior of the hyperbolic plane, and acts as a rotation through angle α where $\cos(\alpha/2) = \mathrm{tr}(g)/2$.
2. If $|\mathrm{tr}(g)| = 2$ then g is the identity or *parabolic*. It fixes no points in the hyperbolic plane, and fixes a unique point at infinity. In the upper half-space model, it is conjugate to a translation $z \to z + 1$.
3. If $|\mathrm{tr}(g)| > 2$ then g is *hyperbolic*. It fixes two unique points at infinity, and acts as a translation along the geodesic joining these points through distance l where $\cosh(l/2) = \mathrm{tr}(g)/2$.

Different models are better for visualizing the action of different isometries. An elliptic isometry is easily visualized in the unit ball model, where the center can be taken to be the origin, and the isometry is realized by an ordinary (Euclidean) rotation. A parabolic

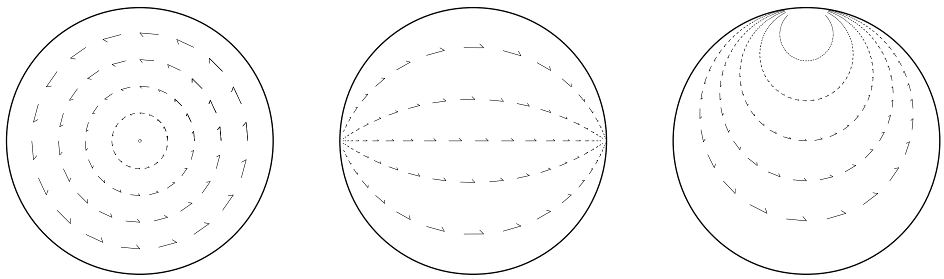

FIGURE 3. Elliptic, hyperbolic and parabolic isometries in the unit disk model.

isometry is visualized in the upper half-space model as a translation. A hyperbolic isometry is visualized in the upper half-plane model as a dilation centered at the origin.

1.3.2. *Quadratic forms and an exceptional isomorphism.* The isometry group of $\mathbb{H}^2$ in the upper half-plane model is naturally isomorphic to the group $\mathrm{PSL}(2,\mathbb{R})$; this expresses the exceptional isomorphism of Lie groups $\mathrm{PSL}(2,\mathbb{R}) = \mathrm{SO}^+(2,1)$. We can see this at the level of Lie algebras by looking at the Killing form. The Lie algebra $\mathfrak{sl}(2,\mathbb{R})$ consists of real 2×2 matrices with trace zero. A basis for the Lie algebra consists of the matrices

$$X := \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad Y := \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \quad H := \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

In this basis, the Lie bracket satisfies $[X,Y] = H$, $[H,X] = 2X$, $[H,Y] = -2Y$. From this and the antisymmetry of Lie bracket, we can express the adjoint action in terms of 3×3 matrices

$$\mathrm{ad}(X) = \begin{pmatrix} 0 & 0 & -2 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \quad \mathrm{ad}(Y) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 2 \\ -1 & 0 & 0 \end{pmatrix}, \quad \mathrm{ad}(H) = \begin{pmatrix} 2 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

The *Killing form* on a Lie algebra is the symmetric bilinear form

$$B(x,y) := \mathrm{trace}(\mathrm{ad}(x)\mathrm{ad}(y))$$

and is invariant under the adjoint action of the group on its Lie algebra. In terms of our given basis, the Killing form B on $\mathfrak{sl}(2,\mathbb{R})$ is given by the symmetric matrix

$$B = \begin{pmatrix} 0 & 4 & 0 \\ 4 & 0 & 0 \\ 0 & 0 & 8 \end{pmatrix}$$

which has two positive eigenvalues and one negative eigenvalue, so the signature is $2,1$ and we obtain a map $\mathrm{SL}(2,\mathbb{R}) \to \mathrm{O}(2,1)$ which factors through the quotient by the center $\pm\mathrm{id}$, and realizes the isomorphism $\mathrm{PSL}(2,\mathbb{R}) = \mathrm{SO}^+(2,1)$.

Another way to see this isomorphism is to think about the action of $\mathrm{SL}(2,\mathbb{R})$ on the space of symmetric quadratic forms in two variables. A symmetric quadratic form $ax^2+2bxy+cy^2$ is represented by a symmetric 2×2 matrix $Q = \left(\begin{smallmatrix} a & b \\ b & c \end{smallmatrix}\right)$ and the group $\mathrm{PSL}(2,\mathbb{R})$ acts on such symmetric forms by $M \cdot Q = M^T Q M$. The *discriminant* of a quadratic form is $\Delta := 4b^2 - 4ac$ which itself is a symmetric quadratic form of signature $(2,1)$. The discriminant is preserved by the $\mathrm{PSL}(2,\mathbb{R})$ action, since it is proportional to $\det(Q)$, and $\det(M) = \det(M^T) = 1$ for $M \in \mathrm{PSL}(2,\mathbb{R})$. The collection of symmetric quadratic forms in two variables with discriminant $-d$ for any positive d is a 2-sheet hyperboloid, and $\mathrm{PSL}(2,\mathbb{R})$ acts on each of these sheets by hyperbolic isometries.

1.3.3. *Unit ball and upper half-space.* In 3 dimensions, the boundary of the upper half-space is identified with $\mathbb{C}$, and the complex projective action of $\mathrm{PSL}(2,\mathbb{C})$ on this boundary extends conformally to the interior. An isometry g might have real trace (in which case it is conjugate into $\mathrm{PSL}(2,\mathbb{R})$ and preserves a totally geodesic 2-plane) or it could be *loxodromic*, in which case it fixes two unique points at infinity, and acts as a "screw motion" along the geodesic joining these points through *complex length* $\ell := l + i\theta$ (i.e. translation length l, rotation through angle θ) where $\cosh(\ell/2) = \mathrm{tr}(g)/2$.

A loxodromic isometry is visualized in the upper half-space model as a dilation centered at the origin together with a rotation about the vertical line through the origin.

1.3.4. *Hermitian forms.* A Hermitian form on $\mathbb{C}^2$ is given by a matrix $Q = \left(\begin{smallmatrix} a & z \\ \bar{z} & b \end{smallmatrix}\right)$ where $a, b \in \mathbb{R}$ and $z \in \mathbb{C}$. Thus, the collection of such forms is a real vector space of dimension 4. The group $\mathrm{PSL}(2,\mathbb{C})$ acts on such forms by $M \cdot Q = \bar{M}^T Q M$ and preserves the discriminant $|z|^2 - ab$ (which, again, is just proportional to $\det(Q)$), a nondegenerate form of signature $(3,1)$. This exhibits the exceptional isomorphism $\mathrm{PSL}(2,\mathbb{C}) = \mathrm{SO}^+(3,1)$.

2. Building hyperbolic manifolds

2.1. **Geometric structures and holonomy.** Fix a Lie (pseudo-)group G and a real analytic manifold X on which G acts effectively.

Definition 2.1. Let M be a manifold. A (G,X) structure is an atlas of charts $\varphi_i : U_i \to X$ on M for which the transition functions are in G. Two such atlases on M are *isomorphic* if they have common refinements which are related by a homeomorphism of M isotopic to the identity.

Let M be a manifold with a (G,X) structure. There is a *developing map* $D : \tilde{M} \to X$ where $\tilde{M}$ denotes the universal cover of M, defined as follows. Pick a basepoint p in M. Then points of $\tilde{M}$ can be identified with homotopy classes rel. endpoints $[\gamma]$ of paths $\gamma : [0,1] \to M$ with $\gamma(0) = p$. If we pick a chart U_0 containing p, there is an *analytic continuation* $\Gamma(\gamma) : [0,1] \to X$ which satisfies $\Gamma(0) = \varphi_0(p)$, and which can be expressed in a neighborhood of each $t \in [0,1]$ in the form $g \circ \varphi_i \circ \gamma$ for some $g \in G$ where g is multiplied by the appropriate transition function when $\gamma(t)$ moves from chart to chart. Then define $D([\gamma]) = \Gamma(\gamma)(1)$.

For each $\alpha \in \pi_1(M,p)$ there is a unique $\rho(\alpha) \in G$ defined by $\Gamma(\alpha * \gamma)(1) = \rho(\alpha)\Gamma(\gamma)(1)$ where $*$ is composition of paths. This defines a homomorphism $\rho : \pi_1(M,p) \to G$ called the *holonomy representation.* A different choice U_k of initial chart containing p would conjugate ρ by $\varphi_k \circ \varphi_0^{-1}$, so really the holonomy representation is well-defined *up to conjugacy.* In the end we obtain a map

$$H : (G,X) \text{ structures on } M/\text{ isomorphism} \to \mathrm{Hom}(\pi_1(M,p),G)/\text{conjugacy}$$

Proposition 2.2 (Thurston [20] Prop. 5.1)**.** *The map H is a local homeomorphism.*

Proof. A conjugacy class of representation $\pi_1(M,p) \to G$ gives rise to an X bundle $X \to E \to M$ over M with a flat G structure. Since it is flat, there is a foliation $\mathcal{F}$ transverse to the fibers, given by the locally constant sections. In this language, a (G,X) structure is nothing but a section $\sigma : M \to E$. Deforming the representation deforms the foliation; since transversality of σ is open, this deformation gives rise to a deformation of the (G,X) structure. □

When X is a complete, simply connected Riemannian manifold and G is its group of isometries, a (G,X) structure on M induces a Riemannian metric. When M is closed, such a metric is necessarily complete, and therefore the developing map $D : \tilde{M} \to X$ is a covering map, which is an isomorphism if $\tilde{M}$ and X are connected. In this case the holonomy representation is discrete and faithful, and $\rho(\pi_1(M))$ acts freely and cocompactly on X.

2.2. **Gluing polyhedra.** To build a hyperbolic structure on a manifold M, it is convenient to decompose M into simple geometric pieces modeled on subsets of $\mathbb{H}^n$ which can be assembled compatibly in limited ways. It is convenient to take for the pieces *convex polyhedra* with totally geodesic faces, which are glued up in isometric pairs (if M is orientable, the isometries are orientation-reversing). A compact polyhedron admits only finitely many isometries, which are determined by how they permute the vertices, but sometimes it is convenient to use noncompact polyhedra, even if M is compact! The reason is that the disadvantage of working with noncompact pieces is greatly outweighed by the advantage of working with pieces whose geometry is described by a small number of moduli.

2.2.1. *Poincaré's polyhedron theorem.* Let's start with a finite collection P_i of n-dimensional hyperbolic polyhedra with totally geodesic faces. For convenience, let's assume the P_i are all compact. A *face pairing* is a choice of (combinatorial) identification of the faces of P_i in pairs which can be realized by an isometric gluing. For compact hyperbolic polyhedra, the isometry is determined by the combinatorics of the pairing.

The result of this gluing is a piecewise-hyperbolic polyhedral complex M. We would like to give necessary and sufficient conditions for this complex to be a hyperbolic manifold. Thus we must check that each point in the complex has a neighborhood which is isometric to an open subset of hyperbolic space. We check this condition on skeleta, starting at the top.

In the interior of the polyhedra P_i there is nothing to check; similarly, the fact that the gluing of faces was done isometrically in pairs means that we have a nice structure on the interior of each codimension 1 face. Suppose ϕ_0 is a codimension 2 face in P_0, so it separates two adjacent codimension 1 faces α_0 and β_0. Now, β_0 is glued to a face α_1 in P_1, identifying ϕ_0 with ϕ_1, which separates α_1 from β_1. Similarly, β_1 is glued to α_2 in P_2, and we obtain a cycle of polyhedra P_i with codimension 2 faces ϕ_i where each is glued to the next successively. Going once round the cycle takes ϕ_0 to itself by an isometry. We want this isometry to be the identity; by compactness this is equivalent to fixing the vertices, in which case ϕ_0 embeds isometrically in M, and there is a hyperbolic structure on the interior of ϕ_0 if and only if the dihedral angles in the P_i along the ϕ_i add up to 2π.

One might think that it is now necessary to impose further more complicated conditions on the faces of codimension 3 and higher, but actually something remarkable happens. Each codimension 3 face in each polyhedron has a linking spherical triangle, and the hyperbolic structures extend to the interior of the codimension 3 faces if these spherical triangles glue up to make a round S^2. A spherical structure on a closed connected 2-manifold R induces a developing map from the universal cover $\tilde{R}$ to S^2. Since R is compact, any Riemannian metric on R is complete, and induces a complete metric on $\tilde{R}$, so the developing map $\tilde{R} \to S^2$ is a covering map, which is automatically an isomorphism. Thus we are reduced to the purely *local* geometric problem of checking that there is a well-defined *spherical* structure on the *link* of every codimension 3 face, plus the purely *topological* problem of checking that the links are all simply-connected. As above, the geometric condition is immediate on the codimension 0 and 1 faces of the spherical triangles, and it follows on the codimension 2 faces by the fact that such faces are the intersections with the codimension 2 faces ϕ_i of the P_i where we have already checked that the (hyperbolic) structure is good.

By induction, on each codimension k face (with $k \geq 3$) we must check that the linking spherical $(k-1)$-simplices glue up to make a round S^{k-1}; equivalently that they are simply-connected, and give rise to a spherical structure. By induction on dimension, it suffices to check this on faces of codimension at most 2, where it follows by examining the codimension 2 faces of the original P_i.

Notice the remarkable fact that we do not even need to check that the complex M is a topological manifold, just that the links of faces of codimension at least 3 are simply-connected. In fact, if we are prepared to work in the category of orbifolds (spaces locally modeled on the quotient of a manifold by a finite group of symmetries) we do not even need to check this.

Notice too that the argument we gave above applies word-for-word to spaces obtained by gluing Euclidean or spherical polyhedra (in fact, the inductive step depends on the proof for spherical polyhedra one dimension lower). Thus we have proved the following theorem of Poincaré:

Theorem 2.3 (Poincaré's polyhedron theorem). *Let $\mathbb{X}^n$ denote n-dimensional hyperbolic, Euclidean, or spherical space, and let P_i be a finite collection of totally geodesic compact polyhedra modeled on $\mathbb{X}^n$. Let M be obtained by gluing the codimension 1 faces of the P_i isometrically in pairs. Suppose for each codimension 2 face ϕ the dihedral angles at ϕ add up to 2π and the composition of the gluing isometries around ϕ are the identity on ϕ. Then M is an orbifold with a complete $\mathbb{X}^n$ structure. If furthermore links in codimension 3 and higher are simply-connected, M is a manifold.*

A more detailed discussion and a careful proof (valid under much more general hypotheses) is given in [6].

The orbifolds that can arise under the hypotheses of Theorem 2.3 have singular locus of codimension at least 3. We should modify the conditions on the gluing in the following ways to obtain arbitrary (compact) orbifolds. Suppose we

(1) allow mirrors on some codimension 1 faces instead of face pairing; and
(2) insist that the link of each codimension 2 face ϕ is a mirror interval of length $\pi/m(\phi)$ or a circle of length $2\pi/m(\phi)$ for some $m(\phi) \in \mathbb{N}$;

then the complex M is a compact orbifold with a complete $\mathbb{X}^n$ structure.

Example 2.4 (Doubling). Let P be a compact 3-dimensional hyperbolic polyhedron with totally geodesic faces, and all dihedral angles of the form π/n for various integers $n \geq 2$. We can give P the structure of a complete hyperbolic orbifold by putting mirrors on all the top dimensional faces, and some finite manifold cover is a closed hyperbolic 3-manifold.

For example, one can obtain a non-compact "super-ideal" regular simplex $\Delta \subset \mathbb{H}^3$ by intersecting a regular simplex in projective space with the interior of the region bounded by a conic (in the Klein model) in such a way that the symmetries of the simplex extend to isometries of hyperbolic space. The dihedral angles between the planes can be chosen to meet at any angle $\alpha < \pi/3$ (the case $\alpha = \pi/3$ corresponds to an inscribed regular simplex — i.e. an equilateral ideal simplex in $\mathbb{H}^3$). For each triple of edges of the simplex meeting at a vertex v outside the conic, there is a (projectively) dual plane in $\mathbb{H}^3$ meeting all three edges perpendicularly. Cut Δ by each of these four planes to obtain a truncated tetrahedron with dihedral angles all equal to α and $\pi/2$. Taking $\alpha = \pi/n$ for $n > 3$ we obtain infinitely many (incommensurable) examples this way.

2.3. **Gluing simplices.** If $\mathbb{X}^n$ in the statement of Theorem 2.3 is hyperbolic space, and we weaken the hypothesis to allow some of the P_i to be noncompact, new phenomena can arise. Rather than pursue this in full generality, we discuss it in the special case that the P_i are ideal simplices, and focus on the case of dimension 2 and 3.

2.3.1. *Ideal triangles and spinning.* An *ideal* polyhedron is one with all its vertices at infinity. In the Klein model, we can think of an (ordinary) Euclidean polyhedron inscribed in the ball. In the upper half-space model, we can put one vertex of an ideal triangle at infinity, then by a Euclidean similarity we can put the other two vertices at 0 and 1. This demonstrates that there is only one ideal triangle up to isometry, making ideal triangles the "ideal" pieces out of which to build hyperbolic surfaces.

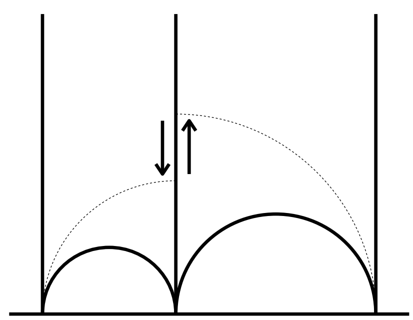

FIGURE 4. When two oriented ideal triangles are glued together along a pair of edges, the gluing is specified by the *shear coordinate* which can be any real number.

When we go to glue two ideal triangles together along an edge a new ambiguity arises. The edges of an ideal triangle are isometric to $\mathbb{R}$, which admits a continuous family of self-isometries. In order to specify the gluing along an edge we must *choose* an isometry from this family. Each edge of an ideal triangle has a canonical midpoint, namely the foot of the perpendicular from the opposite vertex; thus the space of gluings is parameterized by a real-valued *shear coordinate* which measures the (signed) hyperbolic distance that each foot is glued to the right of the other. Notice that to define the shear coordinate we require that our ideal triangles are oriented, and that the gluing is compatible with the orientation, but we do not need to choose an ordering of the edges. In the upper half-space model, we can fix the first triangle to have vertices $-1, 0, \infty$. If after performing the gluing the second triangle has vertices $0, \infty, t$ then the shear coordinate is $\log(t)$. See Figure 4.

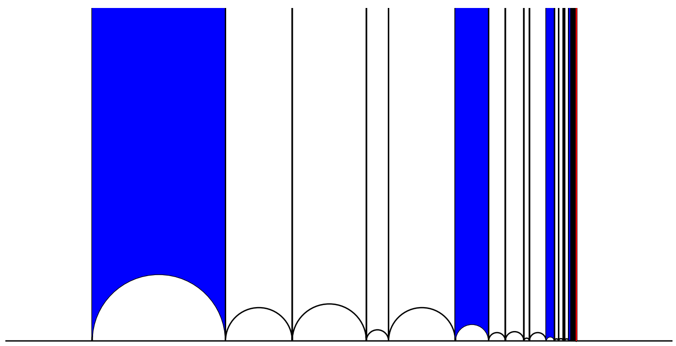

FIGURE 5. Five ideal triangles glued in a loop with hyperbolic holonomy around the cusp. The lifts of one of the triangles are in blue. The triangles accumulate on a "missing" geodesic $\tilde{\gamma}$ (in red). The incomplete structure can be completed by adding the quotient of this missing geodesic by the holonomy around the cusp, which the ends of the ideal triangles all "spin" around.

If finitely many ideal triangles are glued up around an ideal "vertex" with (successive) shear coordinates t_i, the holonomy h of the developing map along a loop around the missing vertex is given (in the upper half-plane model where the ideal vertex is at infinity) by $h : z \to \alpha z + \beta$ for some β depending on the choice of basepoints, and where $\log(\alpha) = \sum t_i$. If $\alpha = 1$ the holonomy is parabolic, and the hyperbolic structure near the omitted vertex is complete. Otherwise, the holonomy is hyperbolic with translation length $\log(\alpha)$, equal to the sum of the shear coordinates on the edges adjacent to the vertex. Under the developing map the lifts of the ideal triangles accumulate on a missing geodesic $\tilde{\gamma}$ which is stabilized by h. The hyperbolic structure on the surface can be completed near the vertex by adding a geodesic $\gamma := \tilde{\gamma}/\langle h \rangle$ of length equal to the absolute value of $\log(\alpha)$, and which the ideal triangles "spin" around, clockwise or anticlockwise depending on whether $\log(\alpha)$ is negative or positive; see Figure 5.

Example 2.5 (Pair of pants). A complete hyperbolic structure on a pair of pants is obtained by doubling an ideal triangle. If instead of doubling the two triangles are glued in the same combinatorial pattern but with arbitrary shears $x, y, z \in \mathbb{R}$ then we obtain a (typically) incomplete structure, that can be completed by adding three boundary geodesics with lengths $|x + y|$, $|x + z|$ and $|y + z|$. The map from shear coordinates to isometry types is a branched cover of degree 8–1, with a twofold ambiguity at each incomplete cusp coming from the choice of whether the ideal triangles spin around the missing geodesic clockwise or anticlockwise.

Example 2.6 (Surface with punctures). Let $\bar{S}$ be a closed oriented surface of genus g with $p > 0$ marked points, and let S be the punctured surface obtained from $\bar{S}$ by removing the marked points. We may triangulate $\bar{S}$ in such a way that the set of vertices is precisely the set of marked points. If t is the number of triangles, then

$$\chi(\bar{S}) = 2 - 2g = t - 3t/2 + p$$

We may obtain a (typically) incomplete hyperbolic structure on S by gluing ideal triangles in the given combinatorial pattern, with one shear parameter for each edge. Thus the space of gluings is $\mathbb{R}^{3t/2} = \mathbb{R}^{6g-6+3p}$. The condition that the structure is complete at a cusp is a linear equation in the shear parameters; thus the space of complete gluings is $\mathbb{R}^{6g-6+2p}$. We shall see that this is a model for the Teichmüller space of S.

2.3.2. *Ideal tetrahedra.* In the upper half-space model, we can move an ideal tetrahedron so that three of its vertices are at $0, 1, \infty$ and its fourth is at $z \in \mathbb{C} - \{0, 1\}$. The number z is called the *simplex parameter*, and is well-defined if we choose a labeling of the vertices. Permuting the vertices induces an action of the symmetric group S_4 on the space of simplex parameters, whose kernel is the Klein 4-group $\mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/2\mathbb{Z}$. Thus the action factors through S_3, whose generators act on simplex parameters by $z \to 1/z$ and $z \to 1/(1-z)$. The orientation-preserving permutations factor through $\mathbb{Z}/3\mathbb{Z}$.

Another way to say it is that the simplex parameter of an ideal tetrahedron is just the (complex) cross-ratio of its vertices. The intersection of an ideal triangle with a horosphere based at a vertex is a Euclidean similarity class of triangle; identifying the Euclidean plane with $\mathbb{C}$, and ordering the vertices somehow, this triangle can be moved so its vertices are at $0, 1, z$. Cyclically permuting the vertices transforms z by

$$z \to \frac{1}{1-z} \to \frac{z-1}{z} \to z$$

Following Neumann–Zagier [16] we sometimes use the abbreviations $z' := (z-1)/z$ and $z'' := 1/(1-z)$. We may associate these parameters to the edges of an (oriented) ideal tetrahedron, and observe that opposite edges (those that don't share a vertex) have the same parameters. If we intersect the ideal tetrahedron with a horosphere centered at a vertex to get a triangle, we inscribe the edge label inside the triangle near the corresponding vertex. See Figure 6.

2.3.3. *Edge and cusp equations.* Suppose X is a 3-dimensional simplicial complex with vertex links all tori, such that the complement of the vertices is homeomorphic to the interior of a compact oriented 3-manifold M. We may try to find a complete hyperbolic structure on M by choosing simplex parameters for the simplices of X that glue together suitably.

When two ideal tetrahedra are glued along faces, there is a unique isometry compatible with any identification of the vertices; thus, once we have chosen the parameters, the precise gluing is forced on us by the combinatorics of X. If we glue finitely many simplices cyclically around an edge e, we must check that we get an honest hyperbolic structure near e. Label each simplex with vertices from 0 to 3 so that 0 is at infinity, so that 01 is the edge e, and vertex 3 of simplex i is glued to vertex 2 of simplex $i+1$. If the simplices (with this ordering) have simplex parameters z_i along the common edge e, then after going once

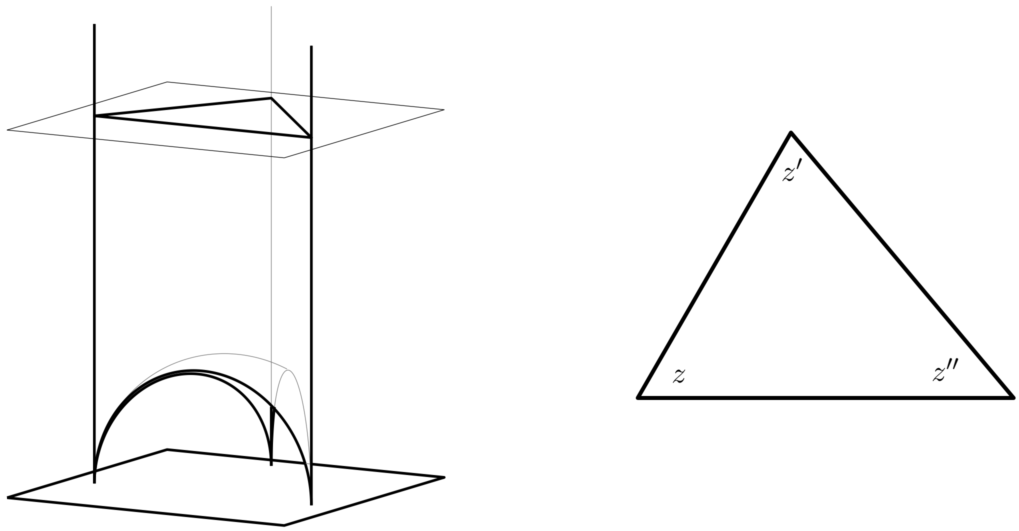

FIGURE 6. An ideal tetrahedron and the simplex parameter associated to the edges.

cyclically around e the holonomy of the gluing is the Möbius transformation $z \to z \cdot \prod z_i$. Thus, one condition is that $\prod z_i = 1$ for each edge e. These are the *edge equations* for the gluing.

The edge equations alone do not ensure that we get a hyperbolic structure on M. If we are gluing oriented simplices, then we want each z_i to have positive imaginary part, so there is a unique value of $\log(z_i)$ whose imaginary part is positive and contained in $(0, \pi)$, and is equal to the dihedral angle of the given simplex along the edge e. To get an honest hyperbolic structure along e, we need that $\sum \log(z_i) = 2\pi i$ for each edge e, with this branch of the logarithm (we may refer to this informally as "the 2π condition").

Let V_X denote the space of solutions to the edge equations. This is a complex affine variety. The subspace $U_X \subset V_X$ where the z_i all have positive imaginary part and the 2π condition is satisfied is an open analytic subvariety, but it is not algebraic. If the edge equations and the 2π condition are satisfied, we obtain a hyperbolic structure on M. However, this hyperbolic structure might be *incomplete*. Complete or not, there is a developing map $D : \tilde{M} \to \mathbb{H}^3$, unique up to isometry, and an associated representation $\rho : \pi_1(M) \to \mathrm{PSL}(2, \mathbb{C})$, unique up to conjugacy.

The map from U_X to representations is nearly injective:

Lemma 2.7. *The map from U_X to $\mathrm{Hom}(\pi_1(M), \mathrm{PSL}(2, \mathbb{C}))$/conjugacy is locally a 2^n to 1 fold branched cover onto its image, where n is the number of cusps of M.*

Proof. Each vertex v of X has a torus link R_v made by gluing ideal vertex links in the simplices of X. The universal cover $\tilde{M}$ may be completed to a simplicial complex $\tilde{X}$ by adding one vertex corresponding to v for each conjugate of the subgroup $\pi_1(R_v)$ in $\pi_1(M)$.

For each solution to the edge equations in U_X there is a developing map $D : \tilde{M} \to \mathbb{H}^3$ that extends to $\tilde{X} \to \mathbb{H}^3 \cup S^2_\infty$, and the image of a vertex $\tilde{v}$ is fixed by the corresponding conjugate of $\rho(\pi_1(R_v))$. Since $\pi_1(R_v)$ is abelian, the group $\rho(\pi_1(R_v))$ is either parabolic (and has a unique common fixed point) or contains hyperbolic elements (and stabilizes a common axis). If $\pi_1(R_v)$ is parabolic, the fixed point is determined uniquely by the group. If $\pi_1(R_v)$ is hyperbolic with a common axis, the endpoints of this axis are determined up to a twofold ambiguity, and are locally determined uniquely. Thus in every case the representation determines the images of the vertices under the developing map, and therefore also the simplex parameters, up to a twofold ambiguity for each non-parabolic cusp. □

The link of an ideal vertex has a canonical complex affine structure, so the surfaces R_v come with developing maps on their universal covers $D : \tilde{R}_v \to \mathbb{C}$ and holonomy representations $\rho : \pi_1(R_v) \to \mathbb{C} \rtimes \mathbb{C}^*$ where $\mathbb{C} \rtimes \mathbb{C}^*$ is the group of complex affine automorphisms of $\mathbb{C}$, with the $\mathbb{C}$ factor acting by translations, and the $\mathbb{C}^*$ factor by dilations. Let $h : \pi_1(R_v) \to \mathbb{C}^*$ be the image of ρ in $\mathbb{C}^*$. Then as in the 2-dimensional case, the hyperbolic structure on M is complete near v if and only if $\rho|\pi_1(R_v)$ is parabolic, or equivalently if and only if h is the trivial representation.

Lemma 2.8 (Geometric point)**.** *Suppose M admits a complete hyperbolic structure obtained by gluing positively oriented ideal tetrahedra. Then the associated solution $p \in U_X \subset V_X$ is unique, and is called the* geometric point*.*

Proof. A complete hyperbolic structure on M, if it exists, is unique up to isometry by the Mostow–Prasad Rigidity Theorem 3.4, to be proved in the sequel. Since the map from U_X to conjugacy classes of representations is injective when all the cusps are parabolic, if the image contains the discrete faithful representation associated to the complete structure, the preimage consists of a single point. □

If we choose meridian m and longitude l generators of $\pi_1(R_v)$, we may express $h(m)$ and $h(l)$ in terms of the simplex parameters z_j. Evidently each of $h(m)$ and $h(l)$ is a product of terms of the form $z_j^{\pm}$, $(z_j')^{\pm}$ or $(z_j'')^{\pm}$. The *cusp equations* set these expressions equal to 1. Thus the edge and cusp equations together define a family of integral algebraic equations in the simplex parameters of quite a simple kind. The solutions of the edge and cusp equations, together with the 2π condition, either pick out a unique point in U_X (the geometric point) or else there is no complete hyperbolic structure on M obtained by gluing positively oriented ideal tetrahedra according to the combinatorics of X.

Example 2.9 (Figure 8 knot complement). Thurston showed that the figure 8 knot complement can be obtained from two regular ideal simplices by a suitable face pairing. See Figure 7.

Give the first tetrahedron the parameter z and the second the parameter w, assigned to specific edges as in Figure 7. The cusp torus R is tiled by eight triangles, the links of the vertices of the two tetrahedra. A fundamental domain is shown in Figure 8.

There are two edges, which impose the edge equations

$$(z'')^2 z' (w'')^2 w' = 1 \text{ and } z^2 z' w^2 w' = 1$$

which are both equivalent to a single equation $z(z-1)w(w-1) = 1$. The space of solutions is a (complex affine) curve of genus 1 and the complete structure corresponds to a point on this curve where $z = w = e^{2\pi i/6}$ and both tetrahedra are regular.

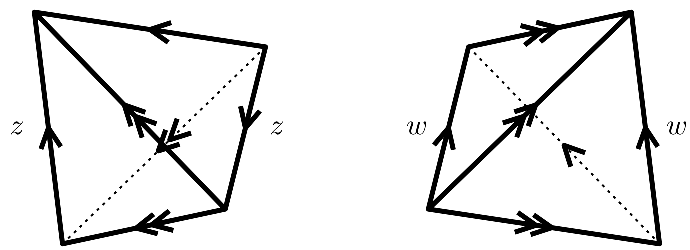

FIGURE 7. Two regular ideal simplices glued with this pairing makes a complete hyperbolic manifold homeomorphic to the complement of the figure 8 knot in S^3.

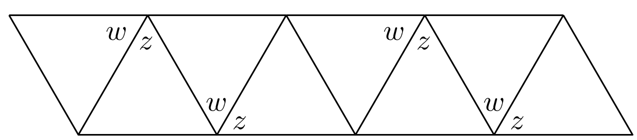

FIGURE 8. The triangulation of the cusp by Euclidean triangles.

2.3.4. *Dimension count of the solution space.* Let's analyze the variety V_X a little more closely. The first thing to do is to estimate its dimension.

Lemma 2.10 (Dimension count)**.** *The complex dimension of V_X is at least as big as the number of cusps of M.*

Proof. Suppose in X there are t simplices and e edges. Taking links of the vertices produces triangulations of the vertex surfaces R_v. Each simplex of X contributes 4 triangles to the R_v, and 12 edges glued in pairs. Each edge of X contributes 2 vertices to the various R_v. Since the R_v are all tori, they satisfy $\chi = 0$ so $4t + 2e = 6t$, or $t = e$. Thus, there are as many edge equations as (ideal) simplex parameters.

However: these equations are *not* independent. For each vertex v we can cut open R_v along the edges of a maximal connected non-separating graph in the 1-skeleton of its triangulation to obtain a polygonal fundamental domain P for R_v made of triangles, and with no interior vertices. Any choice of simplex parameters determines a (possibly non-convex, possibly self-overlapping) complex affine structure on P. Note that P has an even number of edges, since the edges of P are glued in pairs to make R_v.

If P is a polygon with an even number of edges, we can cyclically order the vertices i, and the (oriented) edges e_i so that e_i, e_{i+1} share the common vertex i, and then there is a unique complex affine similarity ϕ_i taking e_{i+1} to e_i by an orientation-reversing map. The complex affine structure on P determines the dilation $w_i \in \mathbb{C}^*$ of ϕ_i. Since the composition of these isometries as we go once around ∂P takes e_1 to itself, and since the number of edges is even, it follows that $\prod w_i = 1$. Each vertex j of P is an edge of M, and under the gluing the vertices of P are partitioned into subsets which are the equivalence classes of some equivalence relation $\sim$. For each equivalence class $[j]$ we see that $\prod_{i\sim j} w_i$ is exactly the edge equation associated to this equivalence class, so it follows that there is exactly one redundancy among the edge equations associated to each cusp and therefore the space of solutions of the edge equations has complex dimension at least as large as the number of cusps. $\square$

It might seem at first glance as though the cusp equations impose two further (complex) conditions for each cusp, but actually it is (generically) true that these equations are dependent. This is because the fundamental group of a torus is abelian, so that $\rho(m)$ and $\rho(l)$ are *commuting* elements of $\mathbb{C} \rtimes \mathbb{C}^*$. Thus, if $h(m) = 1$ and $\rho(m)$ is nontrivial, it follows that we must also have $h(l) = 1$.

2.4. **Hyperbolic Dehn surgery.** Let's specialize now to the case that M has 1-cusp T, and there is a complete hyperbolic structure on M obtained by gluing (positively oriented) ideal tetrahedra whose simplex parameters solve the edge and cusp equations and satisfy the 2π condition. These parameters pick out the geometric point p on the variety V_X. A neighborhood of p gives a 2–1 parameterization of the space of nearby (typically) incomplete structures on M. Developing these structures gives a map from V_X to representations into $\mathrm{PSL}(2, \mathbb{C})$ up to conjugacy.

The restriction of a representation to $\pi_1(T)$ is conjugate into $\mathbb{C} \rtimes \mathbb{C}^*$, though not uniquely; such a choice depends on a choice of a common fixed point at infinity for $\rho(\pi_1(T))$. This choice is determined by the ideal triangulation; thus a point in V_X determines a representation from $\pi_1(T)$ into $\mathbb{C} \rtimes \mathbb{C}^*$, unique up to conjugacy *in* $\mathbb{C} \rtimes \mathbb{C}^*$. In particular, the dilation parts of the holonomy $h(m), h(l) \in \mathbb{C}^*$ are the coordinates of a well-defined map $h : V_X \to (\mathbb{C}^*)^2$.

The geometric point p maps to $(1, 1)$ and in fact it is the unique point in U_X in the fiber over $(1, 1)$, by Lemma 2.8. On the other hand, the map h is not open since near the geometric point $\rho(m)$ and $\rho(l)$ are nontrivial commuting elements, so either $h(m)$ and $h(l)$ are both equal to 1, or neither is equal to 1.

Thus, since V_X has complex dimension at least 1, the image under h of a neighborhood of the geometric point has complex dimension exactly 1; in particular the image of the irreducible component of V_X containing the geometric point is a (complex) plane curve.

Under this assumption (i.e. the existence of an ideal triangulation with all simplices positively oriented) we now prove:

Theorem 2.11 (Thurston's hyperbolic Dehn surgery Theorem [20], 5.8.2)**.** *Let M be a 1-cusped oriented hyperbolic 3-manifold with torus cusp T and meridian and longitude m, l. Then for all but finitely many coprime integers (p, q) there is an incomplete hyperbolic structure on M arbitrarily close to the complete structure, which may be completed to a closed hyperbolic manifold homeomorphic to $M_{p/q}$ (the result of p/q Dehn filling of M along T) by adding a closed geodesic γ which is the core of the added solid torus, and whose length goes to 0 as one of $|p|, |q|$ goes to infinity.*

In particular, for any $\epsilon > 0$ the subset of $M_{p/q}$ where the injectivity radius is less than ϵ converges geometrically to the subset of M where the injectivity radius is less than ϵ.

Proof. As above, we have a 2–1 branched map from points on U_X near the geometric point to representations $\rho : \pi_1(M) \to \mathrm{PSL}(2,\mathbb{C})$ up to conjugacy near the discrete faithful representation ρ_0.

The discrete faithful representation is characterized by the condition that the restriction $\rho_0|\pi_1(T)$ is parabolic. Lift this representation to $\mathrm{SL}(2,\mathbb{C})$, and conjugate it into the subgroup $\mathbb{C} \rtimes \mathbb{C}^*$ of upper triangular matrices in the form

$$\rho_0(m) = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad \rho_0(l) = \begin{pmatrix} 1 & c \\ 0 & 1 \end{pmatrix}$$

where c is a complex number with positive imaginary part. We may think of this as a group of translations of $\mathbb{C}$, and we may choose a fundamental parallelogram with vertices at $0, 1, c, 1 + c$.

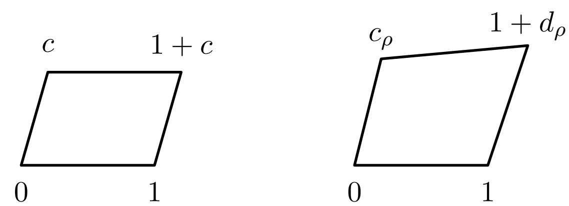

FIGURE 9. A fundamental domain for $\pi_1(T)$ deforms from a parallelogram at the complete structure to a quadrilateral at a nearby structure.

We may lift nearby representations $\rho|\pi_1(T)$ to $\mathrm{SL}(2,\mathbb{C})$. At a nearby ρ the matrices $\rho(m)$ and $\rho(l)$ are nontrivial, nonparabolic and commute, so they fix two points at infinity. The ideal triangulation picks out one of these two points, and therefore determines a conjugate of $\rho|\pi_1(T)$ into $\mathbb{C} \rtimes \mathbb{C}^*$ well-defined up to conjugacy in $\mathbb{C} \rtimes \mathbb{C}^*$. A fundamental quadrilateral domain may be chosen to have vertices $0, 1, c_\rho, 1 + d_\rho$ for some $c_\rho \neq d_\rho$ and $c_\rho, d_\rho \to c$ as $\rho \to \rho_0$. The sides $0, c_\rho$ and $1, 1 + d_\rho$ are paired by $\rho(m)$ and the sides $0, 1$ and $c_\rho, 1 + d_\rho$ are paired by $\rho(l)$. Thus

$$h(m) = \frac{d_\rho}{c_\rho} \text{ and } h(l) = 1 + d_\rho - c_\rho$$

Using the approximations $\log(z) \sim z - 1 \sim (z - z^{-1})/2$ for z close to 1,

$$\frac{\log h(l)}{\log h(m)} \sim \frac{(d_\rho - c_\rho) \cdot (2c_\rho d_\rho)}{d_\rho^2 - c_\rho^2} = \frac{2c_\rho d_\rho}{d_\rho + c_\rho} \sim c$$

Another way to say this is that the image of V_X under h near the geometric point is a smooth plane curve whose tangent line at $(1, 1)$ has slope c.

Since $\log h(l)/\log h(m)$ is almost equal to c near the geometric point, we may find unique real numbers p and q so that $p \log h(m) + q \log h(l) = 2\pi i$, or equivalently

$$p + qc \sim \frac{2\pi i}{\log h(m)}$$

Since $h(m)$ takes values in an open neighborhood of 1 near the geometric point, it follows that we may find representations ρ for which $p \log h(m) + q \log h(l) = 2\pi i$ for all p, q outside a compact neighborhood of $(0, 0)$ in $\mathbb{R}^2$.

For typical p, q the representation ρ is indiscrete. But when p, q are integers and the representation is sufficiently close to the complete structure, then although the hyperbolic structure is incomplete, the holonomy representation *is* discrete, though not faithful since $\rho(m)^p\rho(l)^q = 1$. In this case $\rho(m)$ and $\rho(l)$ stabilize a common geodesic $\tilde{\gamma}$ in $\mathbb{H}^3$ which completes the image of the developing map, and together they generate a group isomorphic to $\mathbb{Z} \oplus \mathbb{Z}/n\mathbb{Z}$ where n is the gcd of p and q. See Figure 10.

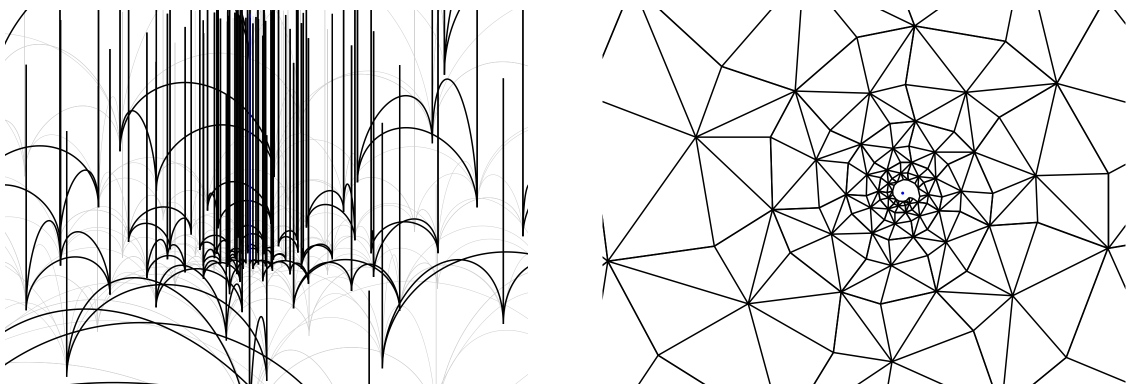

FIGURE 10. Ideal simplices spiraling around a missing geodesic in an unfaithful but discrete representation. The intersection with a horosphere is a collection of Euclidean triangles spiraling around a missing point. This is the 3-dimensional version of Figure 5.

The quotient of $\tilde{\gamma}$ by $\langle \rho(m), \rho(l) \rangle$ is a closed geodesic γ which completes M, giving rise to a hyperbolic structure on the *closed orbifold* whose underlying manifold $M_{p/q}$ is obtained by doing Dehn filling on M along the slope p/q, and which has cone angle $2\pi/n$ along the geodesic γ. If p and q are coprime, then we obtain an honest hyperbolic structure on the closed *manifold* $M_{p/q}$. Each of $\rho(m)$ and $\rho(l)$ acts as a translation of $\tilde{\gamma}$ with translation length $\log|h(m)|$ and $\log|h(l)|$ respectively, and the length of γ is the gcd of these lengths. In particular, the length of γ goes to zero as one of $|p|, |q|$ goes to infinity. The theorem follows. $\square$

Example 2.12 (Generalized Dehn surgery). When p, q are coprime integers for which

$$p \log h(m) + q \log h(l) = t2\pi i \tag{2.1}$$

for some real number t, the holonomy around the curve on the torus with slope p/q is a (typically nontrivial) rotation through angle $t2\pi$. In this case the real parts of $\log(h(m))$ and $\log(h(l))$ generate a rank 1 (and therefore discrete) subgroup of $\mathbb{R}$ so the incomplete hyperbolic structure on M can be (metrically) completed by adding a closed geodesic to obtain a *singular* hyperbolic structure on $M_{p/q}$ which has a cone singularity along the added geodesic, with cone angle $t2\pi$. As we increase t monotonically from 0 to 1, we obtain a one-parameter family of cone manifolds $M(t)$ interpolating between M and $M_{p/q}$. We say these intermediate cone manifolds are obtained by *generalized* hyperbolic Dehn surgery.

Remark 2.13. Theorem 2.11 generalizes to a complete hyperbolic manifold M with any finite number of cusps. One may deform the cusps one at a time, keeping the hyperbolic structure complete in a neighborhood of the other cusps. For each cusp of M, all but finitely many Dehn fillings on this cusp admit complete hyperbolic structures, and the subsets of these filled manifolds where the injectivity radius is bigger than ϵ converge geometrically to the corresponding subset of M.

Remark 2.14. The assumption that M admits an ideal triangulation with all simplices positively oriented is not really necessary; its only role in the proof is to show that the space of incomplete structures on M near the complete structure has complex dimension at least 1. This fact may be proved in a number of ways, and we refer the interested reader to Thurston [20], 5.8.2.

Example 2.15 (Napoleon's 3-manifold). Let L be the 4-component link in Figure 11. Two components of L form a Hopf link H (in red), whose complement is a product $T \times (-1, 1)$. The other two components H' form an unlink (in black) that may be projected to an alternating link on the T. There is an order 2 involution ι of $M := S^3 - L$ that interchanges the two components of H and the two components of H'. The quotient $O := M/\iota$ is an orbifold with two cone geodesics of order 2 and two cusps.

The manifold M may be decomposed into two regular ideal octahedra, each of which may be further decomposed into four ideal tetrahedra meeting along an interior edge running between the two components of H. Let $z_1, \cdots, z_4$ be the simplex parameters associated to one octahedron, and $w_1, \cdots, w_4$ the simplex parameters associated to the other, where the parameters in each case are associated to the interior edge they all share in common. For the complete structure on M every parameter is equal to i.

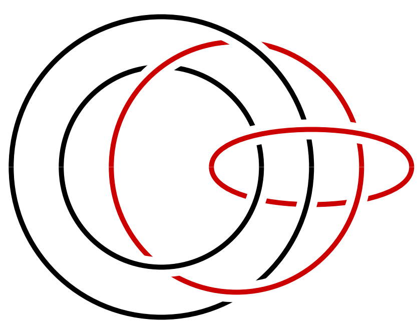

FIGURE 11. Napoleon's 4-component link L in S^3.

The space of solutions of the edge equations is 4 complex dimensional, and the subspace of solutions invariant under the involution ι is 2 complex dimensional, and parameterizes incomplete hyperbolic structures on O. It turns out that this subspace is a product, parameterized by two complex numbers α, β with the following properties:

(1) half of the simplex parameters depend only on α and half depend only on β; and
(2) the holonomy $h(m), h(l)$ at one cusp depends only on α, and the holonomy at the other cusp depends only on β.

Let $O_{p,q,r,s}$ denote the result of p/q filling of one cusp and r/s filling of the other, and let $V_{p,q,r,s}$ denote the volume of $O_{p/q,r/s}$. Then it follows that there are functions f and g so that

$$V_{p,q,r,s} = f(p/q) + g(r/s)$$

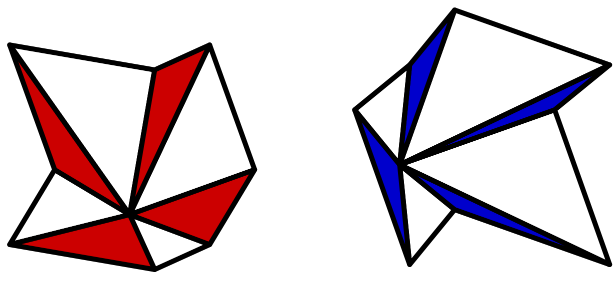

FIGURE 12. Two complex parameters α and β each determine four simplex parameters (red and blue triangles respectively). The white triangles are all isosceles right triangles with a vertex at the corner of a square and a corner at α resp. β.

3. Rigidity and the thick-thin decomposition

3.1. **Mostow rigidity.** The purpose of this section is to prove the following

Theorem 3.1 (Mostow Rigidity Theorem). *Let M, N be closed hyperbolic manifolds of dimension at least 3, and let $f : M \to N$ be a homotopy equivalence. Then f is homotopic to an isometry.*

We prove this theorem following Gromov (rather than giving Mostow's original proof) using the machinery of Gromov norms.

Since hyperbolic manifolds are $K(\pi, 1)$'s, two such manifolds M, N are homotopy equivalent if and only if their fundamental groups are isomorphic. Moreover, outer automorphisms of $\pi_1(M)$ induce self homotopy equivalences of M. Since the group of isometries of a closed Riemannian manifold is a compact Lie group, it follows that $\mathrm{Out}(\pi_1(M))$ is finite whenever M is closed and hyperbolic of dimension at least 3.

3.1.1. *Quasi-isometries.* Let $f : M \to N$ be a homotopy equivalence between closed hyperbolic manifolds, with homotopy inverse $g : N \to M$. We may assume these maps are smooth, and therefore Lipschitz. These lift to Lipschitz maps $\tilde{f} : \tilde{M} \to \tilde{N}$ and $\tilde{g} : \tilde{N} \to \tilde{M}$ between the universal covers (which are both isometric to $\mathbb{H}^n$) whose composition satisfies $d(\tilde{g}\tilde{f}(p), p) \le C$ for some constant C independent of $p \in \tilde{M}$. It follows that $\tilde{f}$ (and likewise $\tilde{g}$) is a *quasi-isometry*; i.e. there exists a constant K so that for all $p, q \in \tilde{M}$ we have

$$\frac{1}{K} d_{\tilde{N}}(\tilde{f}(p), \tilde{f}(q)) - K \le d_{\tilde{M}}(p, q) \le K d_{\tilde{N}}(\tilde{f}(p), \tilde{f}(q)) + K$$

If γ is a geodesic in $\mathbb{H}^n$, we can define a function $\rho : \mathbb{H}^n \to \mathbb{R}^+$ to be the distance to γ. Nearest point projection defines a retraction $\pi : \mathbb{H}^n \to \gamma$. If $S_t(\gamma)$ denotes the level set $\rho = t$, then $d\pi | TS_t$ is strictly contracting, with norm $1/\sinh(t)$. It follows that for every geodesic γ the image $\tilde{f}(\gamma)$ is contained within distance $O(\log(K))$ of some unique geodesic δ, and the map $\tilde{f}$ extends *continuously* (by taking endpoints of γ to endpoints of δ as above) to a *homeomorphism* $\tilde{f}_\infty : S^{n-1}_\infty \to S^{n-1}_\infty$. which intertwines the actions of $\pi_1(M)$ and $\pi_1(N)$ at infinity.

3.1.2. *Gromov norm.* If X is a topological space, the group of real simplicial k-chains $C_k(X;\mathbb{R})$ is not just a real vector space, but a real vector space with a *canonical* basis, consisting of the singular k-simplices $\sigma : \Delta^k \to X$. It makes sense therefore to define an L_p norm on $C_k(X;\mathbb{R})$ for all k, and in particular the L_1 norm which we denote simply $\|\cdot\|$, defined by

$$\|\sum t_i\sigma_i\| = \sum |t_i|$$

for real numbers t_i and singular simplices $\sigma_i : \Delta^k \to X$.

Definition 3.2 (Gromov norm)**.** For a (singular) homology class $\alpha \in H_k(X;\mathbb{R})$, the *Gromov norm* of α, denoted $\|\alpha\|$, is the infimum of $\|z\|$ over all real k-cycles z representing α.

The name Gromov "norm" is misleading, since it could easily be 0 on some nonzero α. In fact, it is not at all obvious that this norm is not identically zero. Note that any continuous map between topological spaces $f : X \to Y$ induces maps $f_* : H_*(X;\mathbb{R}) \to H_*(Y;\mathbb{R})$ which are norm non-increasing. Thus the Gromov norm is invariant under homotopy equivalences. For M a closed, oriented n-manifold, "the" Gromov norm of M is defined to be the norm of the fundamental class; i.e. $\|[M]\|$. It follows that if M and N are homotopy equivalent, they have equal Gromov norms.

The following theorem is key:

Theorem 3.3 (Gromov proportionality)**.** *Let M be a closed, oriented hyperbolic n-manifold where $n \geq 2$. Then*

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

where v_n is the supremum of the volumes of all geodesic n-simplices.

Proof. We first show that $\|[M]\| \geq \text{volume}(M)/v_n$. This inequality will follow if we can show that for any cycle $\sum t_i\sigma_i$ there is a homologous cycle $\sum t_i'\sigma_i'$ where every $\sigma_i' : \Delta_n \to M$ is totally geodesic, and $\sum |t_i| \geq \sum |t_i'|$. In fact, one can make this association functorial, by constructing a chain map $s : C_*(M;\mathbb{R}) \to C_*(M;\mathbb{R})$ taking simplices to geodesic simplices, which is chain homotopic to the identity.

The map s is defined on singular simplices $\sigma : \Delta_n \to M$ as follows. First, lift σ to a map to the universal cover $\tilde{\sigma} : \Delta_n \to \mathbb{H}^n$ where we think of $\mathbb{H}^n$ as the hyperboloid sitting in $\mathbb{R}^{n+1}$. The map $\tilde{\sigma}$ can be straightened to a linear map $\Delta_n \to \mathbb{R}^{n+1}$, and (radially) projected to a totally geodesic simplex in $\mathbb{H}^n$ (this is called the *barycentric* parameterization of a geodesic simplex). Finally, this totally geodesic simplex can be projected back down to M, and the result is $s(\sigma)$. Evidently s is a chain map. Using the linear structure on $\mathbb{R}^{n+1}$ gives a canonical way to interpolate between id and s, and shows that s is chain homotopic to the identity, so induces the identity map on homology. This proves the first inequality.

We next show that $\|[M]\| \le \text{volume}(M)/v_n$, thereby completing the proof. It will suffice to exhibit a cycle $\sum t_i\sigma_i$ representing $[M]$ and with all t_i positive, for which each $\sigma_i(\Delta_n)$ is totally geodesic, with volume arbitrarily close to v_n.

Let Δ denote an isometry class of totally geodesic hyperbolic n-simplex with $|v_n - \text{volume}(\Delta)| < \epsilon/2$. Then it is a fact that for any fixed constant C, and for ϵ sufficiently small, any other totally geodesic simplex Δ' whose vertices are obtained from those of Δ by moving them each a distance less than C, satisfies $|v_n - \text{volume}(\Delta')| < \epsilon$. The group $\text{Isom}(\mathbb{H}^n)$ acts transitively with compact point stabilizers on the space $D(\Delta)$ of isometric maps from Δ to $\mathbb{H}^n$, and we can put an invariant locally finite measure μ on $D(\Delta)$. It is possible to think of a point in $\pi_1(M)\backslash D(\Delta)$ as an isometric map $\Delta \to M$, and to think of the whole space itself with the measure μ as a "measurable" singular n-chain in M, where by convention we parameterize each Δ by the standard simplex with a barycentric parameterization in such a way that the map to M is orientation-preserving. In fact, this space is really a (measurable) *n-cycle*, since for each $\Delta \to M$ and each face ϕ of Δ there is another isometric map $\Delta \to M$ obtained by reflection in ϕ, and the contributions of these two maps to ϕ under the boundary map will cancel. One can in fact develop the theory of Gromov norms for measurable homology, but it is easy enough to approximate this "measurable" chain by an honest geodesic singular chain whose simplices are nearly isometric to Δ.

Choose a basepoint $p \in M$ and let p_1 denote a lift to the universal cover $\tilde{M} = \mathbb{H}^n$. Let E be a compact fundamental domain for M, so that $\mathbb{H}^n$ is tiled by copies gE with $g \in \pi_1(M)$, each containing a single translate gp_1. For the sake of brevity, we denote $p_g := gp_1$. Now, if we denote an $(n+1)$-tuple $(g_0, \cdots, g_n) \in \pi_1(M)^{n+1}$ by $\vec{g}$ for short, we define $c(\vec{g})$ to be the μ-measure of the subset of $D(\Delta)$ consisting of isometric maps $\Delta \to \mathbb{H}^n$ sending the vertex i into g_iE. Furthermore, we let $\sigma_{\vec{g}} : \Delta_n \to \mathbb{H}^n$ denote the singular map sending the standard simplex to the totally geodesic simplex with vertices p_{g_i}. The group $\pi_1(M)$ acts diagonally (from the left) on $\pi_1(M)^{n+1}$, and the projection $\pi \circ \sigma_{\vec{g}}$ is invariant under this action. We can therefore define a *finite* sum

$$z := \sum_{\vec{g} \in \pi_1(M)\backslash \pi_1(M)^{n+1}} c(\vec{g})\pi \circ \sigma_{\vec{g}}$$

which is a geodesic singular chain in $C_n(M;\mathbb{R})$ with all coefficients positive, and for which every simplex has volume at least $v_n - \epsilon$. Just as before z is actually a cycle, and represents a positive multiple of $[M]$. This proves the desired inequality, and the theorem. □

It is a theorem of Haageruup and Munkholm [9] that v_n is equal to the volume of the regular ideal n-simplex, and this is the unique geodesic simplex with volume v_n. So $v_2 = \pi$, $v_3 = 1.014\cdots$ and so on. This is not important for the proof of Theorem 3.3, but it simplifies the proof of Theorem 3.1.

3.1.3. *End of the proof.* If $f : M \to N$ is a homotopy equivalence, it induces an isometry on Gromov norms, and therefore $\text{volume}(M) = \text{volume}(N)$. As in the proof of Theorem 3.3 we can find a geodesic cycle z representing $[M]$ whose simplices are all as close as we like in shape to some fixed Δ of volume arbitrarily close to v_n. The set of vertices of lifts of simplices in the support of z give $(n+1)$-tuples of points in the closed unit ball. Say that a configuration of $(n+1)$ distinct points on S^{n-1}_∞ is *regular* if it is the set of endpoints of a regular ideal n-simplex. By construction, every regular configuration is arbitrarily close to the vertices of some $(n+1)$-tuple in the support of some z. It follows that the map $\tilde{f}_\infty$ must take regular configurations to regular configurations. When $n \ge 3$ there is a unique way to glue two regular n-simplices isometrically along their boundaries, so $\tilde{f}_\infty$ commutes with the (right) action of the group Γ on S^{n-1}_∞ generated by reflections in the side of a regular ideal simplex. Orbits of Γ on S^{n-1}_∞ are dense, so we conclude that $\tilde{f}_\infty$ is *conformal*. Hence the actions of $\pi_1(M)$ and $\pi_1(N)$ are conjugate in $\mathrm{Isom}(\mathbb{H}^n)$ and it follows that M and N are isometric. This completes the proof of Theorem 3.1.

3.1.4. *Maps of nonzero degree.* If $f : M \to N$ is a map between closed oriented hyperbolic manifolds of degree d, Theorem 3.3 and the definition of Gromov norm implies that $\mathrm{volume}(M) \ge d \cdot \mathrm{volume}(N)$, even if M and N have dimension 2. A refinement of Mostow's rigidity theorem due to Thurston says that we have a *strict* inequality $\mathrm{volume}(M) > d \cdot \mathrm{volume}(N)$ unless f is homotopic to a covering map of degree d.

Since $f : M \to N$ is not *a priori* π_1-injective, it is not true that $\tilde{f} : \tilde{M} \to \tilde{N}$ is a quasi-isometry, and there is no reason to expect that it extends continuously to $\tilde{f}_\infty : S^{n-1}_\infty \to S^{n-1}_\infty$.

This can be remedied as follows. If we choose a finite symmetric generating set S for $\pi_1(M)$, it makes sense to define simple random walk on $\pi_1(M)$ with respect to S; i.e. we define a random sequence $g_0, g_1, g_2 \cdots \in \pi_1(M)$ by $g_0 = \mathrm{id}$, and each successive $g_i^{-1} g_{i+1}$ is sampled uniformly and independently from S. Choosing a basepoint $p \in M$ and a lift $\tilde{p} \in \tilde{M}$, we obtain a random walk $g_i(\tilde{p})$ in $\tilde{M}$. Since f has positive degree, $f_*(S)$ generates a subgroup of $\pi_1(N)$ of finite index, and we can define simple random walk on $\pi_1(N)$ with respect to $f_*(S)$ (with the measure obtained by pushing forward the uniform measure on S). A theorem of Furstenberg (which we shall return to in § 5.3.1) says that simple random walks as above converge a.s. to a unique point on the boundary sphere, so we can use this correspondence to define a *measurable* extension of $\tilde{f}$ to $\tilde{f}_\infty : S^{n-1}_\infty \to S^{n-1}_\infty$ conjugating the actions of $\pi_1(M)$ and $\pi_1(N)$. As above, one concludes that if this map does not take regular configurations to regular configurations a.e. then the volume inequality is strict. A measurable map taking regular configurations to regular configurations a.e. turns out to be conformal, and we conclude that f is isometric to a covering map in this case.

3.1.5. *Complete manifolds of finite volume.* A generalization of Mostow Rigidity to complete finite volume manifolds was obtained by Prasad. The statement is as follows:

Theorem 3.4 (Mostow–Prasad Rigidity)**.** *Let M and N be complete finite volume hyperbolic manifolds of dimension at least* 3*. Any proper homotopy equivalence $f : M \to N$ is properly homotopic to an isometry.*

Proof. This can be proved along similar lines to the closed case. A proper homotopy equivalence $f : M \to N$ does not lift *a priori* to a quasi-isometry $\tilde{f} : \tilde{M} \to \tilde{N}$ but with some work one can show that it extends at least to a homeomorphism of boundaries, or alternately Furstenberg's argument shows there is a measurable extension to the sphere at infinity obtained by pushing forward random walk.

Gromov proportionality continues to hold for complete manifolds of finite volume; if one denotes the compact manifold whose interior M has the complete hyperbolic structure by $\bar{M}$, and if $[\bar{M}]$ denotes the fundamental class in $H_n(\bar{M}, \partial\bar{M}; \mathbb{R})$ then there is still an equality $\|[\bar{M}]\| = \text{volume}(\bar{M})/v_n$. However, proving this requires more care. Straightening simplices gives a volume inequality in one direction. Showing the converse — that there are chains with almost all simplices of almost maximal volume — is harder. One elegant argument is due to Kuessner [12]. The notation $M_{(0,\epsilon]}$ for the ϵ-thin part of M (where the injectivity radius is at most ϵ) is explained in § 3.2. For each big ℓ we can find small constants $0 < \epsilon < \epsilon_1$ where $M_{(0,\epsilon_1]}$ is a neighborhood of the end, and $d(\partial M_{(0,\epsilon]}, \partial M_{(0,\epsilon_1]}) > \ell$ (the latter inequality is roughly equivalent to $\epsilon_1/\epsilon > e^\ell$). We can construct, as above, a chain z with support consisting of simplices of volume close to v_n, and with all edges of length close to ℓ, and with all vertices contained in the "thick" part $M_{[\epsilon,\infty)}$. This chain is *not* a cycle, but a face in the support of ∂z is within ℓ of some point in $M_{(0,\epsilon]}$, and is therefore contained in $M_{(0,\epsilon_1]}$. Thus z represents a relative cycle representing a multiple of the fundamental class in $H_n(M, M_{(0,\epsilon_1]}) \cong H_n(\bar{M}, \partial\bar{M})$ where the latter map is induced by a deformation retraction, which induces a chain map of norm 1.

The rest of the proof follows just as in the closed case. □

3.2. **Margulis lemma.** Let M be a complete hyperbolic n-manifold (not necessarily compact). For any $\epsilon > 0$ we define the ϵ-*thin part* of M, denoted $M_{(0,\epsilon]}$, to be the closed subset where the injectivity radius is at most $\epsilon/2$, and the ϵ-*thick part*, denoted $M_{[\epsilon,\infty)}$, to be the closed subset where the injectivity radius is at least $\epsilon/2$. The *Margulis Lemma* is the statement that in each dimension n there is a universal positive constant ϵ_n so that the ϵ_n-thin part of any complete hyperbolic n-manifold has a very simple topology. Explicitly:

Theorem 3.5 (Margulis Lemma). *In each dimension n there is a positive constant ϵ_n so that for any complete hyperbolic n-manifold M, each component of $M_{(0,\epsilon_n]}$ has virtually nilpotent fundamental group. In particular, each component is either a tube — possibly of zero thickness — around an embedded geodesic of length $\leq \epsilon_n$, or a product neighborhood of a cusp.*

3.2.1. *Commutators in Lie groups.* If G is any Lie group, taking commutators defines a smooth map $[\cdot,\cdot] : G \times G \to G$. This map is constant on the factors $G \times \text{id}$ and $\text{id} \times G$, and consequently the derivative is identically zero at $\text{id} \times \text{id}$. Fix a left-invariant Riemannian metric on G and denote $|g| = d(g, \text{id})$. Then there is some ϵ so that if $|g|, |h| < \epsilon$, we have an inequality

$$|[g,h]| \leq \frac{1}{2}\min(|g|,|h|) \tag{3.1}$$

From this we deduce the following lemma:

Lemma 3.6. *For any Lie group G with a left-invariant metric there is an ϵ so that if Γ is a discrete subgroup of G, and Γ_ϵ is the subgroup of Γ generated by elements g with $|g| < \epsilon$, then Γ_ϵ is nilpotent.*

Proof. Because of the identity $[a, bc] = [a, b][b, [a, c]][a, c]$ (valid in any group), to prove that a group is nilpotent it suffices to exhibit an m such that m-fold commutators of the generators are trivial. But if $g_0, \cdots, g_m \in \Gamma$ have $|g_i| < \epsilon$ then

$$|[\cdots[g_0, g_1], g_2], \cdots, g_m]| < 2^{-m}\epsilon$$

Since Γ is discrete, there is some m such that the only $g \in \Gamma$ with $|g| < 2^{-m}\epsilon$ is id. □

Note that whereas ϵ depends only on G, the nilpotence depth m of Γ_ϵ may depend on Γ.

3.2.2. *End of the proof.* Now fix a hyperbolic manifold M and some point $p \in M$. Since M admits a complete hyperbolic structure, $\pi_1(M)$ is a discrete subgroup of $\mathrm{Isom}(\mathbb{H}^n)$.

Define a metric on $\mathrm{Isom}(\mathbb{H}^n)$ by $|g| = d(\tilde{p}, g\tilde{p}) + |\tau(g)|$ where $\tau(g) \in \mathrm{O}(n)$ is the rotation of $T_{\tilde{p}}\mathbb{H}^n$ induced by applying g at $\tilde{p}$ and then parallel transporting back to $\tilde{p}$ along the geodesic from $g\tilde{p}$ to $\tilde{p}$, and $|\cdot|$ is some bi-invariant metric on $\mathrm{O}(n)$. We claim that for any ϵ there is an ϵ' so that if Γ' is the subgroup of $\pi_1(M)$ generated by g with $d(\tilde{p}, g\tilde{p}) \le \epsilon'$, the group Γ' contains with finite index Γ, the subgroup of $\pi_1(M)$ generated by elements with $|g| < \epsilon$. Choosing ϵ as in Lemma 3.6 and taking $\epsilon_n = \epsilon'$ the proof of the first part of Theorem 3.5 will be complete. Let S' denote the set of $g \in \pi_1(M)$ with $d(\tilde{p}, g\tilde{p}) < \epsilon'$ and let S denote the set of $g \in \pi_1(M)$ with $|g| < \epsilon$. Thus $\langle S' \rangle = \Gamma'$ and $\langle S \rangle = \Gamma$.

To prove the claim, write an arbitrary element w of Γ' as a product

$$w = g_1 g_2 g_3 \cdots g_m$$

where each $g_i \in S'$. Now, it is not quite true that τ is a homomorphism from G to $\mathrm{O}(n)$, but the difference between $\tau(gh)$ and $\tau(g)\tau(h)$ is controlled by the curvature tensor, which is quadratic in $d(\tilde{p}, g\tilde{p})$ and $d(\tilde{p}, h\tilde{p})$. Since $\mathrm{O}(n)$ is compact, there is a C depending only on ϵ such that for any C elements of $\mathrm{O}(n)$ there are two with distance at most $\epsilon/8$. Thus we may find distinct indices $i, j \le C$ (assuming $m \ge C$) so that $\tau(g_{i+1} \cdots g_j) < \epsilon/4$. We may furthermore assume that $\epsilon' < \epsilon/4C$ so that the product g of at most C elements of S' has $d(\tilde{p}, g\tilde{p}) < \epsilon/4$, and thus $|g_{i+1} \cdots g_j| < \epsilon/2$. Now, the metric on $\mathrm{Isom}(\mathbb{H}^n)$ is not conjugation invariant, but it is invariant under conjugation by $\mathrm{O}(n)$. Since $\mathrm{O}(n)$ is compact, so we may suppose that the metric on $\mathrm{Isom}(\mathbb{H}^n)$ has the property that $|g^h| < 2|g|$ for $|g| < \epsilon$ and for h sufficiently close to $\mathrm{O}(n)$; i.e. (taking ϵ' small enough) for arbitrary h with $d(\tilde{p}, h\tilde{p}) < C\epsilon'$. Thus we may rewrite

$$g_1 g_2 \cdots g_j = (g_{i+1} \cdots g_j)^{g_1 \cdots g_i} g_1 \cdots g_i$$

where the first term is in S. Inductively, we may express an arbitrary $w \in \Gamma'$ as a product of elements of S times a product of at most $C-1$ elements of S'. Thus Γ has finite index in Γ' as claimed, and we have proved the first part of Margulis' Lemma.

To complete the proof we must analyze the virtually nilpotent discrete torsion-free subgroups of $\mathrm{Isom}(\mathbb{H}^n)$. Consider some component K of $M_{(0,\epsilon_n]}$ with fundamental group Γ. Any two hyperbolic elements with disjoint fixed points at infinity together generate a group which contains free subgroups, by Klein's pingpong lemma. And any two hyperbolic elements with exactly one fixed point in common generate an indiscrete group. So if Γ contains a hyperbolic element g with fixed points $p^\pm$ then every element of Γ must fix both $p^\pm$. Since Γ is torsion-free and discrete, it follows that $\Gamma = \mathbb{Z}$ in this case, and K is a tube around an embedded geodesic.

If Γ contains no hyperbolic elements, then it consists entirely of parabolic elements, which must all have a common fixed point at infinity. In this case K is a product neighborhood of a cusp. This completes the proof.

3.3. **Volumes of hyperbolic manifolds.**

3.3.1. *Gauss-Bonnet theorem.* Gromov proportionality (Theorem 3.3) says that for a closed hyperbolic surface Σ there is an equality

$$\text{area}(\Sigma) = -2\pi\chi(\Sigma)$$

In the sequel it is important to consider surfaces with variable curvature in hyperbolic 3-manifolds. For such surfaces, curvature and topology controls area (and, more importantly, diameter) through the following:

Theorem 3.7 (Gauss-Bonnet)**.** *Let Σ be a closed Riemannian 2-manifold. Then*

$$\int_\Sigma K d\text{area} = 2\pi\chi(\Sigma)$$

where K denotes the Gauss curvature.

Proof. Sectional curvature is tensorial, and therefore on a surface is captured by a 2-form Kdarea which measures the amount of rotation of the tangent space under parallel transport around an infinitesimal parallelogram. By integrating this relationship we see that $\int_\Omega K d\text{area}$ is equal (up to integer multiples of 2π) to the rotation of the tangent space under parallel transport around the oriented boundary $\partial\Omega$ for any domain Ω. Taking $\Omega = \Sigma$ we see that $\int K d\text{area}$ is an integer multiple of 2π, and is therefore independent of the choice of metric (or indeed, the connection). So choose a flat metric with finitely many singularities at each of which there is a cone point. Decomposing into Euclidean triangles whose angles sum to 2π, and using Euler's formula $\chi = F - E + V$ the theorem follows. $\square$

Even if the surface Σ is not smooth everywhere, providing parallel transport makes sense on "enough" curves, it is possible to define curvature as a (signed) *Radon measure* on Σ in such a way that the Gauss-Bonnet theorem is still valid. For example, if Σ is a polyhedral surface made from totally geodesic triangles, there could be atoms of (positive or negative) curvature at the vertices.

3.3.2. *Volumes of ideal simplices.* Recall that an (oriented) ideal simplex, together with a labeling of the vertices, is determined by a complex number $z \in \mathbb{C} - \{0, 1\}$, and permutations of the labels act on the parameter by permuting the values z, $1/(1-z)$ and $(z-1)/z$. Denote the (oriented) volume of an ideal simplex with parameter z by $D(z)$. The function $D(z)$ is single-valued, continuous, and real analytic in $\mathbb{C}$ away from 0 and 1, and evidently satisfies

$$D(z) = D\left(\frac{1}{1-z}\right) = D\left(\frac{z-1}{z}\right), \quad D(z) = -D(1-z) = -D(z^{-1}) \tag{3.2}$$

Five distinct points $0, 1, \infty, z, w$ in $\mathbb{CP}^1$ span five different ideal simplices. If they are oriented in the obvious way as the "boundary" of a degenerate ideal 4-simplex, the sum of their algebraic volumes is zero. Thus there is a *5-term relation*

$$D(z) - D(w) + D\left(\frac{w}{z}\right) - D\left(\frac{1-w}{1-z}\right) + D\left(\frac{1-w^{-1}}{1-z^{-1}}\right) = 0 \tag{3.3}$$

It turns out that D as above is the *Bloch–Wigner dilogarithm*, defined by

$$D(z) := \arg(1-z)\log|z| - \text{Im}\left(\int_0^z \log(1-z)d(\log z)\right) \tag{3.4}$$

One way to discover this is to read a book on special functions, and guess D from the identities that it satisfies. Another method, using the Schläfli formula, will be given in the proof of Proposition 3.11.

A related formula involves the so-called *Lobachevsky function*

$$\Lambda(\theta) := -\int_0^\theta \log|2\sin t|dt \tag{3.5}$$

The volume of an ideal simplex has a very elegant description in terms of Λ:

Proposition 3.8. *If Δ is an ideal simplex with dihedral angles α, β, γ then*

$$\text{volume}(\Delta) = \Lambda(\alpha) + \Lambda(\beta) + \Lambda(\gamma)$$

Note by the way that $\alpha = \arg(z)$, $\beta = \arg((z-1)/z)$ and $\gamma = \arg(1/(1-z))$ up to suitable permutation.

Proof. We compute in the upper half-space model. We put three of the vertices on the unit circle in $\mathbb{C}$ and the fourth at ∞. The three finite vertices a, b, c span a hemispherical triangle whose apex lies above 0 at (Euclidean) height 1. The Euclidean triangle with vertices a, b, c can be subdivided into six right-angled triangles with common vertex at 0 and angles α, β, γ (in pairs). So it suffices to compute the volume of the region σ_α above one of these six triangles, say with angle α, and show it is $\Lambda(\alpha)/2$.

We compute

$$\begin{aligned} \text{volume}(\sigma_\alpha) &= \int_0^{\cos\alpha} dx \int_0^{x\tan\alpha} dy \int_{\sqrt{1-x^2-y^2}}^\infty \frac{dz}{z^3} \\ &= \frac{1}{2}\int_0^{\cos\alpha} dx \int_0^{x\tan\alpha} \frac{dy}{1-x^2-y^2} \\ &= \frac{1}{4}\int_0^{\cos\alpha} \log\frac{\sqrt{1-x^2}\cos\alpha + x\sin\alpha}{\sqrt{1-x^2}\cos\alpha - x\sin\alpha}\frac{dx}{\sqrt{1-x^2}} \end{aligned} \tag{3.6}$$

Doing the substitution $x = \cos t$ gives

$$\begin{aligned} \text{volume}(\sigma_\alpha) &= -\frac{1}{4}\int_{\pi/2}^{\alpha} \log\frac{\sin t\cos\alpha + \cos t\sin\alpha}{\sin t\cos\alpha - \cos t\sin\alpha}dt \\ &= \frac{1}{4}\int_\alpha^{\pi/2} \log\frac{\sin t+\alpha}{\sin t-\alpha}dt \\ &= \frac{1}{4}\int_{2\alpha}^{\alpha+\pi/2} \log|2\sin t|dt - \frac{1}{4}\int_0^{\pi/2-\alpha}\log|2\sin t|dt \\ &= \frac{1}{4}\left(-\Lambda(\alpha+\pi/2) + \Lambda(2\alpha) + \Lambda(\pi/2-\alpha)\right) \end{aligned} \tag{3.7}$$

Now, the angle doubling formula for sin implies the identity

$$\Lambda(2\theta) = 2(\Lambda(\theta) + \Lambda(\theta+\pi/2) - \Lambda(\pi/2))$$

for any θ. Taking $\theta = \pi/2$ gives $2\Lambda(\pi) = \Lambda(\pi)$ so $\Lambda(\pi) = 0$ and we see that Λ is periodic with period π. Since it is evidently odd (because the integrand is even), it follows that

$\Lambda(\pi/2) = 0$. Oddness and π-periodicity give $\Lambda(\alpha + \pi/2) = -\Lambda(\pi/2 - \alpha)$, so Equation 3.7 simplifies to $\text{volume}(\sigma_\alpha) = \Lambda(\alpha)/2$ and the proposition is proved. □

3.3.3. *Schläfli's formula.* Suppose $P(t)$ is a smooth 1-parameter family of hyperbolic n-dimensional polyhedra with a fixed combinatorial type P. For each codimension two face e of P there is a face $e(t)$ of $P(t)$ which has $(n-2)$-dimensional volume $\ell_e(t)$ and dihedral angle $\theta_e(t)$. Let volume(t) denote the n-dimensional volume of $P(t)$. Then there is a remarkable differential formula for the *variation* of volume(t) as a function of t due essentially to Schläfli:

Theorem 3.9 (Schläfli's formula)**.** *With notation as above, there is a differential identity*

$$\frac{d\,\text{volume}(t)}{dt} = -\frac{1}{n-1}\sum_e \ell_e(t)\frac{d\theta_e(t)}{dt} \tag{3.8}$$

Proof. There is a uniform proof that works in all dimensions, but for clarity we will assume $n = 3$. We start by showing that it suffices to reduce to a special case where the computation simplifies. First, since both sides of the formula are additive under decomposition, it suffices to assume P is a simplex. Second, it suffices to prove the formula for finitely many variations whose derivatives span the space of deformations of a simplex. A simplex is cut out by 4 totally geodesic planes, and we consider deformations which keep all but one plane fixed, and move the last plane π by a parabolic motion fixing a point of π at infinity, and with (horocircular) orbits perpendicular to π. The set of such motions spans the space of all deformations, so this is sufficient to prove the theorem.

Fix coordinates in the upper half space so that π is vertical and parallel to the y-z plane, and intersects P in a triangle Δ. Cyclically label the oriented edges of Δ as e_1, e_2, e_3 so that e_1 is contained in the intersection of π with the unit hemisphere centered at the origin (in the x-y plane). The x coordinate is constant on Δ, and we consider a deformation of P obtained by moving Δ by translating it by dx. For this motion, there is a formula

$$\frac{d\,\text{volume}}{dx} = \int_\Delta \frac{dy \wedge dz}{z^3} = \frac{1}{2}\int_{\partial\Delta} \frac{dy}{z^2} \tag{3.9}$$

where the second equality follows by Stokes' theorem.

If θ denotes the dihedral angle along the edge e_1, then $x = \cos(\theta)$ and $z_{\max} = \sin(\theta)$ where $z_{\max}$ is the maximum z coordinate on the geodesic containing e_1. We parameterize the semicircle containing e_1 by angle ϕ so that $y = z_{\max}\cos(\phi)$ and $z = z_{\max}\sin(\phi)$ and observe that the arclength formula gives

$$\ell_{e_1} = \int_{e_1} \frac{dy}{z\sin(\phi)} = \int_{e_1} \frac{dy\, z_{\max}}{z^2} = -\frac{dx}{d\theta}\int_{e_1}\frac{dy}{z^2} \tag{3.10}$$

and there are similar formulae for e_2, e_3. Putting this together with equation 3.9 the formula follows when $n = 3$. Other n follow in essentially the same way. □

One immediate corollary of the Schläfli formula is a new proof of infinitesimal volume rigidity for hyperbolic 3-manifolds. Let M be a closed hyperbolic 3-manifold, and suppose there is some 1-parameter family of deformations of the hyperbolic structure $M(t)$. We can choose some family of fundamental domains $P(t)$ so that volume$(P(t))$ = volume$(M(t))$. Since $P(t)$ can be glued up to form a closed manifold, the edges of P can be partitioned into subsets with the same length whose dihedral angles sum to 2π. Thus Schläfli immediately shows that volume($M(t)$) is constant, recovering a weak version of Gromov proportionality.

Another corollary is a volume inequality for manifolds obtained by Dehn surgery on a cusped manifold. Suppose M is complete finite volume with a cusp, and let $M_{p/q}$ be obtained from M by p/q Dehn surgery (in some coordinates). As in Example 2.12, for all but finitely many p, q there is a 1-parameter family of cone manifolds $M(t)$ for $t \in (0, 1)$ interpolating between M and $M_{p/q}$, which have a singular geodesic where the cone angle is $t2\pi$. We can cut open the $M(t)$ to a polyhedron $P(t)$ in such a way that the singular geodesic is one of the edges of $P(t)$. Then Schläfli's formula says that the derivative of the volume of $M(t)$ is $-\pi\ell(t)$ where $\ell(t)$ is the length of the cone geodesic in $M(t)$. In particular, this derivative is strictly *negative*, so we obtain a strict inequality volume($M_{p/q}$) $<$ volume(M). On the other hand, as p, q converge to infinity, the geometric structures on $M_{p/q}$ converge on compact subsets to that of M, so the volumes must converge. In other words: the map from manifolds to volumes is finite-to-one on $\{M_{p,q}\}$, and the image is a bounded and well-ordered subset of $\mathbb{R}$ of ordinal type ω, whose limit (the supremum, which is not achieved) is equal to volume(M).

In fact, it is possible to estimate the difference in volumes volume(M) $-$ volume($M_{p/q}$) to leading order (for big p, q) in terms of the cusp shape $c := a + bi$. Combining Equation 2.1 with the estimate $\log(\lambda(t)) \sim c\log(\mu(t))$ (where $\mu(t) = h(m)$ and $\lambda(t) = h(l)$) we obtain the formulae

$$\log(\mu(t)) \sim t\frac{2\pi i(p + qa - qbi)}{(p + qa)^2 + (qb)^2}, \quad \log(\lambda(t)) \sim t\frac{2\pi i(p + qa - qbi)(a + bi)}{(p + qa)^2 + (qb)^2} \tag{3.11}$$

The length of the core geodesic $\ell(t)$ is the greatest common "divisor" of the real parts of $\log(\mu(t))$ and $\log(\lambda(t))$, which is

$$\ell(t) \sim \frac{2\pi tb}{(p + qa)^2 + (qb)^2} \tag{3.12}$$

and therefore, by using Schläfli and integrating, we get the following estimate, first obtained by Neumann-Zagier [16] using direct methods:

Proposition 3.10. *Let $M_{p/q}$ be obtained by p/q Dehn surgery on the complete cusped hyperbolic manifold M whose cusp has shape parameter (i.e. ratio of the holonomy of the longitude to the meridian) $a + bi$. Then there is an estimate*

$$\text{volume}(M) - \text{volume}(M_{p/q}) = \frac{\pi^2 b}{(p + qa)^2 + (qb)^2} + O(p^{-4} + q^{-4}) \tag{3.13}$$

The quadratic form $Q(p, q) := ((p + qa)^2 + (qb)^2)/b$ may be given an "intrinsic" definition as the dimensionless quantity which is the length squared of the curve $pm + ql$ on the cusp torus, divided by the area of the torus. If M has more than one cusp and we Dehn fill the cusps independently, the volume contributions from each cusp just add (to leading order).

A further application of Schläfli is to give a derivation of the formula for the volume of an ideal simplex in terms of the Bloch-Wigner dilogarithm. We explain this now.

Proposition 3.11. *If $\Delta(z)$ is an ideal simplex with parameter z then*

$$\text{volume}(\Delta(z)) = \arg(1-z)\log|z| - \text{Im}\left(\int_0^z \log(1-z)d(\log z)\right) \tag{3.14}$$

Proof. In the upper half-space put the four vertices of Δ at 0, 1, z, ∞. Let H_∞ be the horoball consisting of points with Euclidean height T, and let H_0, H_1, H_z be horoballs centered at 0, 1, z with Euclidean height $1/T$. We write $\Delta_v(T) = \Delta \cap \cup_v H_v$ and $P(T)$ for the complement. As $T \to \infty$ the volume of $P(T)$ converges to that of Δ. We will compute $d\text{volume}(\Delta(z))/dz$ by applying Schläfli to compute $d\text{volume}(P(T))/dz$ for fixed T (ignoring the horoball faces) and taking the limit as $T \to \infty$. This is justified, since the horoball faces are extremely small when T is large.

At $T = 1$ these four horoballs have 4 tangencies on the edges $\infty 0$, $\infty 1$, ∞z, 01 and are distance $2\log|z|$ and $2\log|1-z|$ apart along the edges $0z$ and $z1$. For any other value of T the six distances between pairs of horoballs all change by the same constant $2\log T$. But since the dihedral angles of an ideal simplex sum to 2π, changing all the lengths by the same constant contributes 0 to Schläfli.

The dihedral angles along the edges $0z$ and $z1$ are $\arg(1/(1-z))$ and $\arg(z)$ respectively. Taking $T \to \infty$ we obtain by Schläfli a formula for the derivative of volume as a function of z:

$$\frac{d\text{volume}\Delta(z)}{dz} = \log|z|\frac{d\arg(1-z)}{dz} - \log|1-z|\frac{d\arg z}{dz} \tag{3.15}$$

Since $\text{volume}(\Delta(0)) = 0$ we can simply integrate this from 0 to z (along a contour with $\text{Im}z$ positive). Integrating the first term by parts gives

$$\text{volume}(\Delta(z)) = \log|z|\arg(1-z) - \int_0^z \arg(1-z)\frac{d\log|z|}{dz} + \log|1-z|\frac{d\arg z}{dz} \tag{3.16}$$

and the two terms under the integral sum to $\text{Im}\left(\log(1-z)d(\log z)/dz\right)$, completing the proof. ∎

Example 3.12 (Special values). The Bloch–Wigner dilogarithm D is related to the classical dilogarithm $\text{Li}_2(z) := \sum_{k=1}^\infty z^k/k^2$ by the relation $D(z) = \text{ImLi}_2(z) + \arg(1-z)\log|z|$. In this form particular values of D may be related to special values of Dirichlet zeta functions.

Taking $z = e^{i\pi/3}$ gives a regular ideal simplex. Taking $z = i$ ...

3.3.4. *Thurston-Jorgenson theorem.* The following remarkable theorem follows almost formally from what we have done so far:

Theorem 3.13 (Thurston, Jorgenson)**.** *The set of volumes of finite volume complete hyperbolic 3-manifolds is a closed, well-ordered subset of $\mathbb{R}$ or order type ω^ω.*

Proof. From Thurston's hyperbolic Dehn surgery Theorem 2.11, Mostow's Rigidity Theorem 3.1 (and its strengthening due to Prasad), and the Neumann-Zagier volume formula in Proposition 3.10, the theorem will follow once we show that for any positive number V, there is a *finite set* of finite volume cusped manifolds $M_1, \cdots, M_n$ (where n and the M_j depend on V of course) such that *every* finite volume complete hyperbolic manifold M with $\text{volume}(M) \le V$ is obtained by Dehn filling some subset of the cusps of one of the M_j.

But this in turn follows immediately from the Margulis Lemma; i.e. Theorem 3.5. If V is fixed, there are only finitely many possibilities for the topology of the thick part $M_{[\epsilon,\infty)}$ for any M with volume$(M) \le V$. But the complement of the thick part consists of cusps and embedded solid torus tubes around short geodesics. The claim and the theorem follow. □

Notice that the method of proof and Proposition 3.10 actually imply that the map from manifolds to volumes is finite (though unbounded) to one. The statement of the theorem requires some interpretation. It says first that the set of volumes are ordered as

$$v_0 < v_1 < \cdots < v_\omega < v_{\omega+1} < \cdots < v_{2\omega} < \cdots < v_{3\omega} < \cdots < v_{\omega^2} < \cdots < v_\kappa < \cdots$$

where each κ is an infinite ordinal which is a "polynomial" in ω; i.e.

$$\kappa = a_0 + a_1\omega + a_2\omega^2 + \cdots + a_n\omega^n$$

where all the a_i are non-negative integers, and a_n is positive. Said in words, this theorem says there is a smallest volume, a second smallest volume, and so on; then a first "limit" volume — i.e. a smallest volume which is a nontrivial limit (from below) of smaller volumes, and a first "limit of limit volumes", and so on to all finite orders.

Every number of the form volume(M) where M is finite volume but noncompact with j cusps is a limit volume; i.e. it is of the form v_κ for $\kappa = a_k\omega^k + \cdots + a_n\omega^n$ with a_k nonzero, for some $k \ge j$.

Example 3.14 (Small volume orientable manifolds). The Thurston-Jorgenson theorem holds with exactly the same statement (and essentially the same proof) if one restricts attention to volumes of *orientable* finite volume hyperbolic 3-manifolds. Several values of (orientable) volumes v_κ^+ associated to "simple" ordinals κ and the manifolds they correspond to are known by now, including:

- $v_0^+ \sim 0.942707\cdots$ is uniquely the volume of the *Weeks manifold*; i.e. $(5/1, 5/2)$ filling on the Whitehead link complement (Gabai-Meyerhoff-Milley, [8]);
- $v_\omega^+ \sim 2.02988\cdots$ is the volume of the figure 8 knot complement and of its "sister"; i.e. $(5/1)$ filling on one component of the Whitehead link complement (Cao-Meyerhoff, [4]); and
- $v_{\omega^2}^+ \sim 3.66386\cdots$ is the volume of the Whitehead link complement and of the $(-2, 3, 8)$ pretzel link complement (Agol, [1]).

Note that v_ω^+ is twice the volume of the regular ideal simplex, and $v_{\omega^2}^+$ is the volume of the regular ideal octahedron, and in fact the associated minimal volume manifolds can be obtained by gluing up these polyhedra.

Example 3.15 (Negative volumes). We may like to extend the Thurston–Jorgenson theorem to encompass volumes of spherical manifolds, and it remains true if we take the convention that the 'hyperbolic volume' of a spherical manifold M is the negative of its spherical volume. Any spherical manifold M is finitely covered by S^3, so in our convention volume$(M) = -2\pi^2/d$ where d is the degree of the covering projection (equivalently, the cardinality of $\pi_1(M)$). This set of volumes is still well-ordered, of order type ω! In fact, we can put together the volumes of all spherical, hyperbolic and Euclidean 3-manifolds (with the convention that Euclidean 3-manifolds have volume 0) in $\mathbb{R}$ where they form a closed well-ordered set of order type ω^ω. The map from manifolds to volumes is finite-to-one, and the unique minimum (at $-2\pi^2$) is achieved by S^3.

4. Quasiconformal maps and Teichmüller theory

In this section we set up the analytic foundations necessary to study the deformation theory of infinite volume complete hyperbolic structures on 3-manifolds. This is an interesting subject in its own right, and is a key ingredient in the proof of Thurston's hyperbolization theorem for Haken 3-manifolds, and the classification of Kleinian groups. A basic reference for this section is Ahlfors [2] or Hubbard [11].

4.1. **Complex analysis.**

Complex analysis is a vast subject, and we do not treat it here. The purpose of this section is to state some well-known theorems in complex analysis which we shall generalize to the quasiconformal setting in the next few sections. We use the term *conformal* to mean holomorphic and locally injective.

The following three theorems are standard:

Theorem 4.1 (Montel's Theorem)**.** *Let $\Omega \subset \mathbb{C}$ be an open domain, and let $f_n : \Omega \to \mathbb{C}$ be a family of locally bounded holomorphic functions. Then f_n has a subsequence which converges uniformly on compact subsets of Ω to some holomorphic limit f. Furthermore, if the f_n are conformal, then f is either conformal or constant.*

Theorem 4.2 (Riemann Mapping Theorem)**.** *Let $\Omega \subset \mathbb{C}$ be a proper open simply-connected domain. Then there is a conformal isomorphism $f : \Omega \to \mathbb{D}$ unique up to post-composition with a conformal automorphism of $\mathbb{D}$.*

Theorem 4.3 (Conformal automorphisms and isometries)**.** *For any three distinct points $z_1, z_2, z_3 \in \hat{\mathbb{C}}$ there is a unique conformal automorphism of $\hat{\mathbb{C}}$ taking these points (in order) to $0, 1, \infty$. Every conformal automorphism of $\hat{\mathbb{C}}$ extends continuously to a unique isometry of $\mathbb{H}^3$ and conversely.*

4.1.1. *Tensors on Riemann surfaces.*

A Riemann surface S with local holomorphic coordinate z is also a smooth real 2-dimensional manifold with local smooth coordinates x and y. Let TS and T^*S denote the tangent and cotangent space, thought of as real 2-dimensional vector bundles over S. The 1-forms dx and dy are a local basis for smooth sections of T^*S (as a module over the ring of smooth real-valued functions on S). We may complexify this bundle $T^*_{\mathbb{C}}S := T^*S \otimes_{\mathbb{R}} \mathbb{C}$ and then choose a different basis $dz := dx + idy$ and $d\bar{z} := dx - idy$ for the space of smooth sections of $T^*_{\mathbb{C}}S$ (now as a module over the ring of smooth complex-valued functions on S).

Dual to dz and $d\bar{z}$ locally are the complex valued smooth vector fields

$$\frac{\partial}{\partial z} := \frac{1}{2}\left(\frac{\partial}{\partial x} - i\frac{\partial}{\partial y}\right) \text{ and } \frac{\partial}{\partial \bar{z}} := \frac{1}{2}\left(\frac{\partial}{\partial x} + i\frac{\partial}{\partial y}\right)$$

These are a basis for sections of $T_{\mathbb{C}}S := TS \otimes_{\mathbb{R}} \mathbb{C}$. Ordinary vector fields on smooth (real) manifolds are linear differential operators on functions, and so are their complexifications. Thus $\partial f/\partial z$ and $\partial f/\partial \bar{z}$ make sense for any complex-valued smooth function f on S. These expressions are usually abbreviated f_z and $f_{\bar{z}}$ respectively. Note that the 1-forms $f_z dz$ and $f_{\bar{z}} d\bar{z}$ make sense as smooth sections of $T^*_{\mathbb{C}}S$ independently of the choice of local holomorphic function z.

Theorem 4.4 (Differential characterization of conformal maps)**.** *Let S be a Riemann surface and let $\Omega \subset S$ be an open domain with local holomorphic coordinate z. A smooth* *map $f : \Omega \to \mathbb{C}$ is holomorphic if and only if $f_{\bar{z}}$ is identically zero, and is conformal if and only if it is holomorphic and locally injective.*

The real tangent space TS may be thought of as a complex line bundle by letting i act by $i \cdot \partial/\partial x = \partial/\partial y$ and $i \cdot \partial/\partial y = -\partial/\partial x$. This is isomorphic as a complex line bundle to the span of $\partial/\partial z$ in $T_{\mathbb{C}}S$. Thus any expression of the form $f(z)\partial/\partial z$ with f a smooth complex valued function may be interpreted as an 'honest' vector field (in the sense of differential topology) on the real 2-manifold S.

The complexified cotangent bundle $T^*_{\mathbb{C}}S$ splits as a sum of complex line bundles $T^{1,0} \oplus T^{0,1}$ which are the spans of dz and $d\bar{z}$ respectively. This splitting is independent of the choice of local holomorphic parameter z, and for any smooth complex valued function f it induces the decomposition $df = f_z dz + f_{\bar{z}} d\bar{z}$.

4.1.2. *Moduli and extremal length.* A *quadrilateral* is a closed topological disk $Q \subset \mathbb{C}$ with four points $q_1, \cdots, q_4 \in \partial Q$ in cyclic order. We say informally that the points decompose Q into four *sides*: the bottom, right, top, left sides are the arcs of ∂Q running from q_i to q_{i+1} mod 4 for $i = 1, 2, 3, 4$ respectively.

By the Riemann mapping theorem there is a unique *normalized rectangle* R — i.e. a rectangle of the form $R := \{z = x + iy \text{ such that } 0 \le x \le 1, 0 \le y \le m\}$ — and a conformal isomorphism from the interior of Q to the interior of R taking q_1, q_2, q_3, q_4 to $0, 1, 1 + im, im$ in order. The real number $m \in \mathbb{R}^+$ is the *modulus of* Q (i.e. the height divided by the width) and is denoted $m(Q)$. Permuting the vertices cyclically changes m to $1/m$.

To see that the modulus is well defined, suppose we have two normalized rectangles R and R' which are both conformally isomorphic to Q. Then they are conformally isomorphic to each other by a conformal isomorphism $\phi : R \to R'$ taking vertices to vertices (and in particular fixing the vertices 0 and 1). But now we may extend ϕ repeatedly by Schwarz reflection to an isomorphism from $\mathbb{C}$ to $\mathbb{C}$ fixing 0 and 1. Such an isomorphism is necessarily the identity.

Since modulus is defined purely in terms of the conformal geometry of Q, it is a conformal invariant. Conversely, we have the following theorem:

Theorem 4.5 (Geometric characterization of conformal maps)**.** *An orientation-preserving homeomorphism $f : \Omega \to \Omega'$ between domains in $\mathbb{C}$ is conformal if and only if it preserves the modulus of every quadrilateral.*

See Ahlfors [2], Theorem 1.

To actually compute or estimate modulus one uses the method of *extremal length*.

Definition 4.6 (Extremal length)**.** Let $\Omega \subset \mathbb{C}$ be an open domain, and let Γ be a collection of properly embedded paths in Ω. For any Borel measurable non-negative function ρ on Ω and any $\gamma \in \Gamma$ we may define

$$L_\rho(\gamma) := \int_\gamma \rho |dz| \text{ and } A_\rho := \int_\Omega \rho^2 |dz|^2$$

Define $L_\rho(\Gamma)$ to be the infimum of $L_\rho(\gamma)$ over all $\gamma \in \Gamma$. The *extremal length* of Γ, denoted $E(\Gamma)$, is defined to be the supremum of the ratio $L_\rho(\Gamma)^2/A_\rho$ over all ρ with $0 < A_\rho < \infty$.

Actually one ought to insist that the paths in Γ are locally rectifiable, but we ignore this point; for details see [2]. From the definition follows the monotonicity of E: if $\Gamma \subset \Gamma'$ then $E(\Gamma) \geq E(\Gamma')$.

Example 4.7 (Modulus of a quadrilateral). Now, let Q be a quadrilateral and let Γ be the set of all proper paths in Q that run from the left side to the right side. We claim that $E(\Gamma) = 1/m(Q)$.

To see this, suppose $f : Q \to R$ is the continuous extension of the conformal isomorphism between the interior of Q and the interior of a rectangle as above. The image $\Gamma_R := f(\Gamma)$ is precisely the set of proper paths in R running from the bottom to the top edge. If ρ_R is any Borel measurable non-negative function on R, then $\rho := |f'|\rho_R$ is Borel measurable and non-negative on Q, and by change of variables, $E(\Gamma) = E(\Gamma_R)$.

On the other hand, we shall show $E(\Gamma_R) = 1/m$. Taking $\rho = 1$ we have $L_\rho(\gamma) = 1$, realized only on straight horizontal paths, and $A_\rho = m$. Thus $E(\Gamma_R) \geq 1/m$. On the other hand, for any ρ, let's let $\ell = L_\rho(\Gamma_R)$ and define $\gamma_t \in \Gamma_R$ to be the horizontal line with imaginary part $t \in [0, m]$. Then $\ell \leq \int_{\gamma_t} \rho|dz|$ for all t, so by Cauchy–Schwartz

$$m\ell \leq \int_0^m \int_{\gamma_t} \rho|dz|dt \leq \left(\int_R \rho^2|dz|^2 \int_R |dz|^2\right)^{1/2} = (mA_\rho)^{1/2}$$

Hence $A_\rho \geq m\ell^2$ so that $L_\rho(\Gamma_R)^2/A_\rho \leq \ell^2/m\ell^2 = 1/m$. Since ρ was arbitrary, $E(\Gamma_R) \leq 1/m$ and therefore $E(\Gamma) = E(\Gamma_R) = 1/m$.

As a corollary, if Γ are the set of all proper paths in Q that run from the bottom side to the top side then $E(\Gamma) = m(Q)$. One may remember this formula and its relation to extremal length by the mnemonic 'height squared over area equals modulus'.

Example 4.8 (Modulus of an annulus). If $A \subset \mathbb{C}$ is an open annulus then by taking logarithms we may exhibit the universal cover $\tilde{A}$ as a simply-connected open subset of $\mathbb{C}$. If $\tilde{A}$ is all of $\mathbb{C}$ then A is conformally isomorphic to $\mathbb{C}^*$. Otherwise $\tilde{A}$ is isomorphic to the unit disk $\mathbb{D}$ and the deck group of the cover is either a parabolic or hyperbolic isometry. In particular, A is either conformally isomorphic to a punctured disk, or to the region contained between two concentric round circles in $\mathbb{C}$ one of radius 1 and one of radius $r > 1$ (we call this a *round annulus*). The *modulus of* A, denoted $m(A)$, is the number $\log r/2\pi$ in this case, or ∞ if A is conformally equivalent to $\mathbb{C}^*$ or a punctured disk. If Γ is the set of all proper paths in A running between the two boundary components (the *cuffs* of A), then arguing as in Example 4.7 gives $E(\Gamma) = m(A)$ ('height squared over area equals modulus'). Conversely, if Γ is the set of all homotopically essential loops in A then $E(\Gamma) = 1/m(A)$.

Remark 4.9. Some authors use $\log r$ (or even worse, r) in place of $\log r/2\pi$ as the definition of the modulus of an annulus.

Example 4.10 (Inclusion of annuli). Note if $A \subset A'$ where the inclusion is a homotopy equivalence, and Γ, Γ' are systems of paths in A and A' as above, then each $\gamma \subset \Gamma$ is a subset of Γ' and therefore by extending any ρ on A by zero on A' we see $m(A') \geq m(A)$ (and in fact the inequality is strict unless $A = A'$). In fact, if A_i is any countable collection of annuli whose interiors are disjointly embedded in A' by a homotopy equivalence, then $m(A') \geq \sum m(A_i)$ with equality if and only if the A_i are (after a conformal isomorphism) all round concentric annuli whose union is the round annulus A'.

Example 4.11 (Inclusion of quadrilaterals). Let A be an annulus and let Γ be the set of embedded paths running between cuffs. Any $\gamma \in \Gamma$ decomposes A into a quadrilateral Q with left and right sides identified. One may compute the modulus of Q from the subset $\Gamma' \subset \Gamma$ consisting of paths that do not cross γ (since these are exactly the paths that run from the bottom to the top side of Q). Since $\Gamma' \subset \Gamma$ we have $m(Q) = E(\Gamma') \geq E(\Gamma) = m(A)$. A more careful analysis of the equality case of Cauchy–Schwartz shows that $m(Q) = m(A)$ if and only if γ maps to a straight radial arc under a conformal isomorphism from A to a round annulus. The same argument shows that if Q_i are a collection of quadrilaterals with disjoint interior in A whose bottom and top sides respectively are contained in the two cuffs of A, then $\sum 1/m(Q_i) \leq 1/m(A)$ with equality if and only if (after a conformal isomorphism) the Q_i all have left and right edges which are radial straight arcs and whose union is the round annulus A.

Example 4.12 (Annuli in a torus). Every conformal torus T may be obtained from a normalized rectangle R with modulus $1/m$ by gluing left and right sides by a translation (building in this way a Euclidean annulus A with modulus $1/m$) and then gluing the top to the bottom cuffs by a rotation. The cuffs of A glue up to a curve in T whose isotopy class we call the meridian. Suppose γ is any longitude of T, and let Γ be the set of curves on T in the isotopy class of γ. We claim $E(\Gamma) \geq 1/m$ with equality if and only if T is the rectangular torus. To see this, just choose ρ to be the Euclidean metric, and then observe that the area is $1/m$ and the length of every curve in Γ is $\geq 1/m$. Since E is a supremum over all ρ, this gives the lower bound. If $\gamma \in \Gamma$ is any curve, then if we decompose T along γ into an annulus A' the curves $\Gamma' \subset \Gamma$ that do not cross γ are all the essential loops in A'. Thus $1/m(A') = E(\Gamma') \geq E(\Gamma) \geq 1/m$ so $m(A') \leq m$. If we decompose A' further into annuli A_i then as in Example 4.10 we have $\sum m(A_i) \leq m(A') \leq m$.

4.2. **Quasiconformal maps.** Let S be a Riemann surface and $\Omega \subset S$ a domain in S with local holomorphic coordinate z. If $f : \Omega \to \mathbb{C}$ is a smooth function, its Jacobian J_f (as a smooth map between real 2-manifolds) satisfies $J_f = |f_z|^2 - |f_{\bar{z}}|^2$, so where f is locally an orientation-preserving diffeomorphism we have $|f_z|^2 - |f_{\bar{z}}|^2 > 0$ or equivalently $\mu := f_{\bar{z}}/f_z$ has $|\mu| < 1$.

The 1-forms $f_z dz$ and $f_{\bar{z}} d\bar{z}$ are independent of the holomorphic coordinate z, and their ratio is called the *Beltrami differential*:

$$\mu_f = \mu_f(z)\frac{d\bar{z}}{dz} := \frac{f_{\bar{z}}d\bar{z}}{f_z dz} \tag{4.1}$$

This is a smooth section of the complex line bundle $(T^{1,0})^* \otimes T^{0,1}$ which we might write informally as $T^{-1,1}$. We often suppress the f in the subscript of μ when it is clear from context. Notice that the absolute value of a Beltrami differential is well-defined, independent of the choice of (holomorphic) coordinate z, but the argument is not.

A conformal map f has $\mu = 0$, and sends infinitesimal circles to infinitesimal circles. If f is an orientation-preserving diffeomorphism, it sends infinitesimal circles to infinitesimal *ellipses*. The ratio K of the length of the major to the minor axis of these infinitesimal image ellipses is

$$K(z) = \frac{1 + |\mu(z)|}{1 - |\mu(z)|} \tag{4.2}$$

The preimages of the major axes defines an (unoriented) distribution of lines in the tangent space of Ω, at least where $\mu \neq 0$. This is called a *line field*, and it is the covariant avatar of the argument of μ.

A locally injective smooth map f is K-*quasiconformal* in $\Omega \subset S$ if $K := \sup_{z\in\Omega} K(z)$ is finite. The number K is the (maximal) *dilatation* of f in the domain Ω.

4.2.1. *The meaning of K.* If E and F are complex vector spaces of dimension 1, we may denote by $L^+_\mathbb{R}$ the space of oriented real-linear isomorphisms $E_\mathbb{R} \to F_\mathbb{R}$ and by $L_\mathbb{C} \subset L^+_\mathbb{R}$ the space of complex isomorphisms $E \to F$. The coset space $L^+_\mathbb{R}/L_\mathbb{C}$ is isomorphic to the coset space $\mathrm{SL}(2,\mathbb{R})/\mathrm{SO}(2,\mathbb{R})$, which may be identified with $\mathbb{H}^2$, with $L_\mathbb{C}$ corresponding to a marked point in $\mathbb{H}^2$. If $E = T_z\mathbb{C}$ and $F = T_{f(z)}\mathbb{C}$ then $df|z$ determines a point in $L^+_\mathbb{R}/L_\mathbb{C}$, and $\log K(z)$ is the hyperbolic distance of this point from the basepoint.

4.2.2. *Analytic definition.* For various reasons it is important to work with a generalization of the definition of quasiconformal for maps f which are not necessarily smooth.

A real-valued function on an interval $f : I \to \mathbb{R}$ is said to be *absolutely continuous* if for every positive number ϵ there is a positive δ so that for all finite sets of pairwise disjoint open intervals (x_k, y_k) in I satisfying $\sum(y_k - x_k) < \delta$ there is an inequality $\sum |f(y_k) - f(x_k)| < \epsilon$. A function f is absolutely continuous if and only if it is of the form $f(x) = \nu((-\infty, x])$ for some measure ν absolutely continuous (in the sense of measure) with respect to Lebesgue measure. Thus an absolutely continuous function is differentiable almost everywhere.

A function $f : \Omega \to \mathbb{C}$ for a domain $\Omega \subset \mathbb{C}$ is said to be *absolutely continuous on lines* (ACL for short) if the restriction of its real and imaginary parts to almost all horizontal and vertical lines in Ω are absolutely continuous.

Definition 4.13 (Analytic definition of quasiconformal maps)**.** An orientation-preserving homeomorphism $f : \Omega \to \mathbb{C}$ is *quasiconformal* if it satisfies

(1) f is ACL on Ω; and
(2) there exists a k with $0 \le k < 1$ such that $|f_{\bar z}| \le k|f_z|$ a.e. on Ω.

Let $K = (1 + k)/(1 - k)$ for the infimal such k. Then we say f is K-quasiconformal, and that K is the dilatation of f.

Remark 4.14. Some authors, e.g. Hubbard [11] work with distributional partial derivatives, and in place of the ACL condition, imposes the condition that the distributional partial derivatives are locally square integrable (i.e. they are in L^{loc}_2). The equivalence of these two conditions, given an estimate of the form $|f_{\bar z}| \le k|f_z|$, is [2], Lemma 2. Actually, we are rather sympathetic to Hubbard's perspective, since in our view the key geometric property of quasiconformal maps is that they are absolutely continuous with respect to *area* — i.e. null sets are mapped to null sets; and the area of the image may always be obtained by integrating the (distributional) Jacobian ([2], Theorem 3). However, since the definition in terms of ACL is 'standard', and in the sequel we almost always work with the geometric definition of a quasiconformal map (see § 4.2.3) anyway, this point is somewhat moot.

4.2.3. *Geometric definition.* Recall the definition of a quadrilateral and its modulus.

Definition 4.15 (Geometric definition of quasiconformal maps)**.** An orientation-preserving homeomorphism $f : \Omega \to \mathbb{C}$ is *quasiconformal* if there is some constant $K \ge 1$ such that the modulus $m(f(Q)) \le K \cdot m(Q)$ for all quadrilaterals Q in Ω.

We now have two definitions of a quasiconformal homeomorphism — an analytic one and a geometric one. If one takes the analytic definition, the following theorem is the analog of Theorem 4.5:

Theorem 4.16 (Equivalence of definitions)**.** *The analytic and geometric definitions of quasiconformal maps are equivalent. Thus an orientation-preserving homeomorphism in a domain is K-quasiconformal (in the sense of Definition 4.13) if and only if it multiplies the modulus of every quadrilateral by at most K.*

One direction of this theorem (analytic definition implies geometric definition) may be proved via extremal length, since one readily estimates that $E(\Gamma)/K \le E(f(\Gamma)) \le KE(\Gamma)$ for any collection of curves Γ in a domain Ω. For the other direction, see [2] pp. 24–32.

We often abbreviate the expression "K-quasiconformal" by "K-qc" in the sequel. From the definitions one sees that the inverse of a K-qc map is K-qc; that the property of being K-qc is conformally invariant; and that the composition of a K_1 and a K_2-qc map is K_1K_2-qc.

Example 4.17 (Modulus of annuli)*.* By the inequality in Example 4.8 we may conclude that if $f : \Omega \to \mathbb{C}$ is K-quasiconformal, and $A \subset \Omega$ is an annulus, then $m(f(A)) \le K \cdot m(A)$.

Example 4.18 (Quasicircles)*.* A *quasicircle* is a Jordan curve in $\hat{\mathbb{C}}$ that is the image of a (round) circle under a quasiconformal map. It is called a K*-quasicircle* if it is the image under a K-qc map.

Ahlfors gave several geometric characterizations of quasicircles. We discuss one of these characterizations without proof.

Definition 4.19*.* A Jordan curve γ in $\mathbb{C}$ has *bounded turning* if there is a constant C such that if z_1, z_2 are chosen on γ, and z_3 is on the component of $\gamma - \cup z_i$ of least diameter, then

$$|z_1 - z_3| + |z_2 - z_3| \le C|z_1 - z_2| \tag{4.3}$$

One way to think of this condition is that it says the curve does not make "detours" that are large compared to the distance between the endpoints. Since this is a scale-invariant property, it is invariant under conformal automorphisms.

Proposition 4.20. *A Jordan curve is a quasicircle if and only if it has bounded turning.*

4.2.4. *Equicontinuity.* The main advantage of working with this more analytically complicated class of transformations (rather than just working e.g. with smooth quasiconformal maps) is that they satisfy a suitable *equicontinuity* property which allows one to take limits.

Example 4.21 (Quasiconformal invariance of $\mathbb{C}$)*.* Let $f : \mathbb{C} \to \mathbb{C}$ be an injective K-quasiconformal map onto the image. Then $f(\mathbb{C}) = \mathbb{C}$. For otherwise, we could compose f with a conformal map so that the image was contained properly inside the unit disk $\mathbb{D}$. But now we obtain a contradiction. For, if $A_r \subset \mathbb{C}$ is the annulus $A_r := \{z : 1 \le |z| \le r\}$ then $m(A_r) = \log r/2\pi \to \infty$ and therefore $m(f(A_r)) \ge m(A_r)/K \to \infty$ too. However, each $f(A_r)$ includes by a homotopy equivalence into the annulus $B := \mathbb{D} - f(\mathbb{D})$ and therefore by monotonicity of extremal length (Example 4.10) $m(f(A_r)) \le m(B) < \infty$.

A similar argument proves the following analog of Montel's Theorem 4.1:

Theorem 4.22 (Montel's Theorem for quasiconformal maps)**.** *Let $\Omega \subset \mathbb{C}$ be an open domain, and let $f_n : \Omega \to \mathbb{C}$ be a family of locally bounded K-quasiconformal homeomorphisms. Then f_n has a subsequence which converges uniformly on compact subsets of Ω to a limit which is either K-quasiconformal or constant.*

For proofs of Theorem 4.16 and Theorem 4.22 see Ahlfors [2].

4.2.5. *Measurable Riemann Mapping Theorem.* The most important property of quasiconformal maps is a generalization of the Riemann mapping theorem which promises the existence of quasiconformal homeomorphisms with prescribed Beltrami differential. The theorem, due essentially to Morrey [15], is as follows:

Theorem 4.23 (Existence of quasiconformal homeomorphism)**.** *For every measurable Beltrami differential $\mu := \mu(z)d\bar{z}/dz$ on $\hat{\mathbb{C}}$ with $\operatorname{ess\,sup}_z |\mu(z)| < 1$, there is a quasiconformal homeomorphism $f : \hat{\mathbb{C}} \to \hat{\mathbb{C}}$ with dilatation equal to μ a.e. Moreover, f is unique if we further impose that f fixes 0, 1 and ∞.*

We sometimes denote the f associated to μ promised by the theorem by f^μ. Ahlfors-Bers extended Theorem 4.23 to show that f depends holomorphically on μ, where the complex structure on the space of Beltrami differentials is inherited as a subspace of a complex Banach space.

We will not prove this theorem here, but remark that it is easy to directly write down a "conformal class" of metric on $\hat{\mathbb{C}}$ associated to μ, namely any metric of the form $g(z) := \gamma(z)|dz + \mu(z)d\bar{z}|^2$, where $\gamma(z) > 0$ is a measurable, real-valued function of z. If one can then find isothermal coordinates for this new metric, and applies the Uniformization Theorem to the Riemann surface so obtained, the theorem is proved. When the metric is smooth, finding isothermal coordinates is relatively easy (the real analytic case was proved by Gauss); the general case can be derived from the smooth case by approximation and equicontinuity (i.e. Theorem 4.22).

4.2.6. *Quasiconformal maps and quasi-isometries.* There is an intimate and very useful relationship between quasi-isometries of hyperbolic space and quasiconformal maps at infinity.

Theorem 4.24 (Quasiconformal maps extend)**.** *Every quasi-isometry $F : \mathbb{H}^3 \to \mathbb{H}^3$ extends continuously to a quasiconformal homeomorphism of the boundary $f : S^2_\infty \to S^2_\infty$. Conversely, every quasiconformal homeomorphism of a sphere arises this way.*

Proof. Quasi-isometries take geodesics to quasi-geodesics, which are a bounded distance from genuine geodesics. The bound depends on the constant K of quasi-isometry. If γ is a geodesic, let δ be the geodesic obtained by straightening $F(\gamma)$. Now, suppose π is a totally geodesic plane perpendicular to γ at some point. We claim that there is a constant C depending only on K so that the projection of $F(\pi)$ to δ has diameter at most C. This implies in particular that the circle at infinity of π maps to a topological circle which is the core of an annulus of bounded modulus. Taking limits, the continuous extension f takes small round circles to topological circles wedged between round circles of comparable radius, and therefore f is quasiconformal.

To see the claim, consider a geodesic γ' in π intersecting γ, and let δ' be the geodesic obtained by straightening $F(\gamma')$. The image $F(\pi)$ is a bounded distance from the union of such δ', so it suffices to show that the projection of δ' to δ has bounded diameter. But this is equivalent to the condition that there is a bound on the length of segments in δ and δ' which stay within constant distance of each other, which follows immediately from the quasi-isometry property by uniform properness of the distance function on the pair γ, γ'.

Conversely, let f be a quasiconformal homeomorphism of S^2_∞. Douady and Earle define the following *conformal barycenter extension* of f, as follows. For each point $p \in \mathbb{H}^3$, let ν_p denote the visual measure on S^2_∞ as seen from p. Then ν_p pushes forward to the probability measure $f_*\nu_p$ on S^2_∞. Define a vector field V_p on $\mathbb{H}^3$ as follows. For each $q \in \mathbb{H}^3$ identify the unit tangent sphere $U_q\mathbb{H}^3$ with S^2_∞ by the exponential map. Then define

$$V_p(q) = \int_{U_q\mathbb{H}^3} z \, d(f_*\nu_p)(z) \tag{4.4}$$

Heuristically, the point q is "pulled" towards each point at infinity with an intensity proportional to the measure $f_*\nu_p$, and these pulls combine to define the flow.

Douady and Earle [5] show that there is a unique point in $\mathbb{H}^3$ at which the vector field V_p vanishes, called the *barycenter* of the measure. Then defining $F(p)$ to be equal to the barycenter of $f_*\nu_p$ we obtain the desired extension. The precompactness of the space of K-quasiconformal homeomorphisms for fixed K, and the uniqueness of the barycenter, together formally imply that the map F is a quasi-isometry (with a constant that can be estimated in principle from K). □

Remark 4.25. There are easier ways to obtain a quasi-isometric extension of a quasiconformal map. The Douady-Earle extension is canonical and conformally invariant, which are very useful properties. However it is important to note that the extension F is *not* typically a homeomorphism.

Another method to obtain an extension which has the additional property of being a homeomorphism is to associate to f its Beltrami differential μ, and think of f as the time 1 map of a 1-parameter flow f_t obtained by the Measurable Riemann Mapping Theorem (Theorem 4.23) from the family of Beltrami differentials $t\mu$ for $t \in [0,1]$. The derivative of the barycenter extensions of the f_t defines a (time-dependent) flow on $\mathbb{H}^3$, and we obtain an extension F of f by integrating this flow. See e.g. McMullen [14] B.4 for details.

4.3. **Teichmüller space.** Let F be an oriented topological surface, either closed or obtained from a closed surface by removing finitely many points.

Definition 4.26. A *marked hyperbolic structure* on F is a complete finite area oriented hyperbolic surface S together with a homotopy class of orientation-preserving homeomorphism $f : F \to S$. The *Teichmüller space of* F, denoted $\mathcal{T}(F)$, is the space of marked hyperbolic structures on F.

Thus a point in Teichmüller space is an equivalence class of pairs (f, S) where $(f_1, S_1) \sim (f_2, S_2)$ if and only if there is an isometry $\phi : S_1 \to S_2$ such that ϕf_1 is homotopic to f_2.

Actually for now we have only defined $\mathcal{T}(F)$ as a set. We may topologize it in several ways, for instance by putting coordinates on it.

Example 4.27 (Trace coordinates). A marked hyperbolic structure on F is the same thing as a discrete faithful representation $\rho : \pi_1(F) \to \mathrm{PSL}(2,\mathbb{R})$ taking boundary elements to parabolic transformations, up to conjugacy. For each conjugacy class $g \in \pi_1(F)$ the square of the trace $\mathrm{tr}^2\rho(g)$ is well-defined. Knowledge of $\mathrm{tr}^2\rho(g_j)$ for finitely many specific elements $g_j \in \pi_1(F)$ determines the others, and since a discrete faithful representation is irreducible, the (squares of the) traces determine the representation up to conjugacy. Thus we may embed $\mathcal{T}(F)$ in a finite dimensional Euclidean space whose coordinates are the squares of the traces of the g_j.

Example 4.28 (Shear coordinates). Suppose F has at least one puncture. Choose a (combinatorial) ideal triangulation of F with every vertex of every triangle at a puncture. We saw in Example 2.6 that if F has genus g with p punctures, the space of shear coordinates giving rise to a complete hyperbolic structure is $\mathbb{R}^{6g-6+2p}$. Conversely, as in Lemma 2.7, a complete finite area hyperbolic structure gives rise to a representation whose peripheral elements are parabolic. Two triangles sharing an edge determine four conjugacy classes of peripheral elements in $\pi_1(F)$, and under such a representation the unique fixed points of the corresponding parabolic elements determine four points in S^1 whose cross ratio determines the shear coordinate of the edge. Thus the map from shear coordinates to representations is a bijection, and $\mathcal{T}(F) = \mathbb{R}^{6g-6+2p}$ in this case.

Example 4.29 (Fenchel–Nielsen coordinates). Suppose F is closed of genus g. We may decompose F along a system of $3g-3$ disjoint essential simple closed curves δ_i into $2g-2$ pairs of pants. In any marked hyperbolic structure $f : F \to S$ the curves $f(\delta_i)$ are isotopic to geodesics γ_i that decompose S into pairs of pants with totally geodesic boundary.

A hyperbolic pair of pants is determined up to isometry by the lengths of the three boundary cuffs. These lengths must agree when the pants are glued together in S, so they give $3g-3$ parameters. The gluing along each pair of cuffs determines a further parameter, the relative twist; although on the face of it, this twist parameter seems to be defined in $\mathbb{R}/\ell\mathbb{Z}$ where ℓ is the length of the cusp, rotating once around the cuff changes the marking by composition with a Dehn twist, so these parameters are really different as marked hyperbolic structures. In the end one obtains $3g-3$ lengths $\ell_i \in \mathbb{R}^+$ and (after fixing a basepoint somehow) $3g-3$ twist parameters $\alpha_i \in \mathbb{R}$ so that $\mathcal{T}(F)$ is homeomorphic to $\mathbb{R}^{6g-6}$ in this case.

In terms of these parameters we may define a symplectic form $\omega := \sum d\alpha_i \wedge d\ell_i$ on $\mathcal{T}(F)$, which a priori seems to depend on a choice of curve system. However, Wolpert [24] showed that ω is well-defined independent of the choice of curve system, and is intrinsic to $\mathcal{T}(F)$.

4.3.1. *Measured Foliations.* One of the most important set of coordinates for applications to 3-manifolds are *length coordinates*, and their connection to measured foliations and measured laminations. This theory was worked out by Thurston [21] and it would be impossible to do more here than to give definitions and statements of the most important results. For a more careful treatment and for details, see [20] and [7].

For every essential simple closed curve δ on F and every $(f,S) \in \mathcal{T}(F)$ the isotopy class of $f(\delta)$ is represented by a unique geodesic in S, with length ℓ_δ. We may think of ℓ_δ as a function on $\mathcal{T}(F)$, and in this way define a map from $\mathcal{T}(F)$ to $\mathbb{R}^{\mathcal{S}}$ where $\mathcal{S}$ denotes the set of all essential simple closed curves. The image in $\mathbb{RP}^{\mathcal{S}}$ is relatively bounded, and its closure defines a compactification of Teichmüller space. This is easy to see: $3g-3+p$ suitable curves cut up F into pairs of pants (where some cuffs might be punctures if F has punctures); two sets of $3g-3+p$ curves may be chosen which intersect in minimal position and together *bind* the surface — i.e. cut it up so that every complementary region is a finite polygon or a once-punctured finite polygon. The lengths of these $6g-6+2p$ simple curves determine the lengths of every other one, and in fact every ℓ_δ is a Lipschitz function of these $6g-6+2p$ coordinates. What is not easy to see is the topology of this compactification. To state the result we must first introduce some terminology.

Definition 4.30 (Measured singular foliations)**.** A *singular foliation* $\mathcal{F}$ on a surface F is a foliation with 1-dimensional leaves with singularities locally modeled on a branched cover of degree $n/2$ ($n \geq 3$) of the foliation of the plane by horizontal lines. A *transverse measure* is a nonsingular measure μ of full support on transversals that is holonomy invariant. A *measured foliation* is a pair $(\mathcal{F}, \mu)$ where $\mathcal{F}$ is a singular foliation and μ is a transverse measure.

In charts where the foliation looks like a product, the local leaf space is a (transverse) interval; a transverse measure is just a measure on these local leaf spaces which is compatible on the overlap of product charts. Technically, in the theory of foliations, this is what is usually referred to as an *invariant transverse measure* (some important measures on transversals are not holonomy invariant) but the terminology is standard in this setting. For more on foliations and transverse measures in general, see Chapter 4.

If $\mathcal{F}$ is a singular foliation, a singularity of order $n > 3$ may be 'split open' into a union of singularities of order n_i where $\sum(n_i - 2) = n - 2$, and the foliation may be perturbed to a new foliation $\mathcal{F}'$ so that the new singularities are joined up by a tree of segments. A transverse measure on $\mathcal{F}$ determines one on $\mathcal{F}'$ and vice versa, by restricting to transversals away from the singularities. This operation generates an equivalence relation on isotopy classes of singular measured foliations called *Whitehead equivalence*.

If δ is an isotopy class of essential simple closed curve on F, and $(\mathcal{F}, \mu)$ is a measured foliation, we may find a representative of δ that meets $\mathcal{F}$ *essentially*. This means either that we isotop δ into a closed leaf of $\mathcal{F}$, or that we isotop it to be transverse everywhere. Any two representatives of δ transverse to $\mathcal{F}$ are isotopic through transverse representatives, so the measure $\mu(\delta)$ of a transverse representative is well-defined, and defines a function ℓ_δ from measured foliations to non-negative real numbers. This function is well-defined on Whitehead equivalence classes of measured foliations, and lets us make the following definition:

Definition 4.31 ($\mathcal{MF}$)**.** For a surface F with $\chi(F) < 0$ we let $\mathcal{MF}(F)$ denote the space of Whitehead equivalence classes of measured foliations, topologized as a subset of $\mathbb{R}^{\mathcal{S}}$ with coordinates ℓ_δ.

The image of $\mathcal{MF}(F)$ in $\mathbb{R}^{\mathcal{S}}$ is a cone: $\mathbb{R}^+$ acts on $\mathcal{MF}(F)$ by multiplying the transverse measure. The quotient by the orbits of this action are *projective measured foliations* $\mathcal{PMF}(F)$ which embeds homeomorphically in $\mathbb{RP}^{\mathcal{S}}$.

The main theorem about measured foliations is:

Theorem 4.32 (Thurston compactification)**.** *The closure of $\mathcal{T}(F)$ in $\mathbb{RP}^{\mathcal{S}}$ is homeomorphic to a closed ball of dimension $6g-6+2p$. The interior is precisely $\mathcal{T}(F)$ and the boundary is $\mathcal{PMF}(F)$, which consequently is homeomorphic to a sphere of dimension $6g-7+2p$.*

We shall not prove this theorem here (see instead e.g. [21] or [7]). However in the next couple of sections we shall at least see why $\mathcal{MF}(F)$ and $\mathcal{PMF}(F)$ are manifolds.

4.3.2. *Measured laminations.* If F is a closed hyperbolic surface, the universal cover $\mathbb{H}^2$ is compactified by a circle at infinity S^1_∞. In the proof of Mostow rigidity, we saw that any homeomorphism between closed hyperbolic surfaces $f: F \to F'$ lifts to a quasi-isometry between their universal covers and extends to a homeomorphism from S^1_∞ to itself. Lifts of isotopic homeomorphisms extend to the *same* homeomorphism between their circles at infinity, and this homeomorphism topologically conjugates the actions of π_1 (unlike the case of hyperbolic manifolds of dimension 3 or greater, where this conjugacy is actually conformal).

Definition 4.33 (Geodesic lamination)**.** A *geodesic lamination* on a hyperbolic surface F is a closed union of complete disjointly embedded geodesics.

Some of these geodesics might be closed, and some might be infinite in both directions. Another way to think of a lamination is as a foliation of a closed subset of F. In this interpretation, the geodesics are the leaves of the foliation.

Let λ be a geodesic lamination on a hyperbolic surface F. In the universal cover each lift of each geodesic limits to a pair of points (i.e. an S^0) in S^1_∞, and no two of these S^0s link in S^1_∞ (though they may share a point in common). The data of this collection of S^0s is well-defined without reference to a choice of hyperbolic structure, and therefore makes sense for a surface F which is merely hyperbolizable:

Definition 4.34 (Abstract lamination)**.** An *abstract geodesic lamination* on a hyperbolizable surface F is a closed, nonempty, nonlinking, $\pi_1(F)$-invariant collection of S^0s in S^1.

Denote the set of abstract geodesic laminations on a hyperbolizable surface F by $\mathcal{L}(F)$. Notice that the definition of an abstract lamination only depends on the topology of F. If λ is an abstract lamination on F, and $(f,R) \in \mathcal{T}(F)$ then $f: F \to R$ induces a $\pi_1(F)$-equivariant homeomorphism from S^1 to $S^1_\infty(R)$. Each S^0 in λ determines an S^0 in $S^1_\infty(R)$ that spans a unique geodesic in the universal cover $\tilde{R}$. This collection of geodesics is closed, nonintersecting and $\pi_1(F) = \pi_1(R)$ invariant, so it covers an honest geodesic lamination in R.

A geodesic lamination in a hyperbolic surface lifts to a closed subset of the projective tangent bundle, and the set of geodesic laminations may be topologized as a compact subset of the space of closed subsets of the projective tangent bundle with the Hausdorff topology. This set is in bijection with $\mathcal{L}(F)$, and the topology is independent of the choice of hyperbolic structure. Thus $\mathcal{L}(F)$ may be thought of as a topological space.

If F is a hyperbolic surface and λ is a geodesic lamination on F, a transversal σ to λ is *tight* (or *efficient*) if it does not cobound a bigon with any leaf of λ.

Definition 4.35 (Measured lamination)**.** If F is a hyperbolic surface and λ is a geodesic lamination on F, an (invariant) transverse measure for λ is a measure μ on tight transversals that is holonomy invariant.

In this definition we allow measures with atoms and/or without full support. If $f : F \to F'$ is a homeomorphism taking $f : \lambda \to \lambda'$ we may push forward a measure μ on λ to a measure μ' on λ'. Thus it makes sense to talk about measures on abstract laminations.

If δ is an essential simple closed curve on F, there is a function ℓ_δ from measured laminations to non-negative real numbers whose value on a measured lamination (λ, μ) is the μ-measure of a tight representative of δ (note that this does not depend on a choice of hyperbolic structure). This function cannot separate different laminations whose measures are supported on the same sublamination, so we make the following definition:

Definition 4.36 ($\mathcal{ML}$)**.** For F a surface with $\chi(F) < 0$ we define $\mathcal{ML}(F)$ to be the space of laminations with transverse measures of full support, topologized as a subset of $\mathbb{R}^{\mathcal{S}}$ with coordinates ℓ_δ.

As with foliations, $\mathcal{ML}(F)$ is a cone which projects to $\mathcal{PML}(F)$ in $\mathbb{RP}^{\mathcal{S}}$.

Remark 4.37. The topologies on $\mathcal{L}(F)$ and $\mathcal{ML}(F)$ are quite different. In the Hausdorff topology $\mathcal{L}(F)$ is totally disconnected: any 1-parameter family of geodesics on a hyperbolic surface of finite area must contain geodesics with self-intersections. On the other hand, in the length topology $\mathcal{ML}(F)$ is path-connected, and is even a manifold.

There is a subset of $\mathcal{L}(F)$ consisting of laminations that admit measures of full support and this maps continuously and bijectively but not homeomorphically to the quotient space of $\mathcal{ML}(F)$ obtained by forgetting the measure.

4.3.3. *Straightening.* Now suppose F is a hyperbolic surface and $(\mathcal{F}, \mu)$ is a measured foliation. One may put (in many ways) a singular Euclidean metric on F associated to $(\mathcal{F}, \mu)$ with excess cone angle $(n-2)\pi$ at each n-prong singularity, for which the leaves are parallel (Euclidean) straight lines, and the transverse measure is length of orthogonal transversals. This metric is non-positively curved, and segments contained in leaves of $\mathcal{F}$ are distance minimizing curves in their relative homotopy class. It follows that in the universal cover, the (nonsingular) leaves of $\tilde{\mathcal{F}}$ are (uniform) quasigeodesics, which limit to S^0s in S^1_∞. This system of S^0s is unlinked and π_1-invariant, and its closure contains S^0s associated to quasigeodesics contained in lifts of singular leaves of $\mathcal{F}$.

It therefore determines an abstract lamination on F which may be realized by a geodesic lamination λ in the hyperbolic structure on F. We say this lamination is obtained from $\mathcal{F}$ by *straightening*. This operation takes leaves of $\mathcal{F}$ to leaves of λ, though not necessarily injectively.

If τ is a transversal to $\mathcal{F}$ we may obtain an efficient transversal τ' to λ with endpoints on the 'corresponding' leaves. We may define a measure μ' on λ by $\mu'(\tau') = \mu(\tau)$. Thus straightening makes sense as a map from $\mathcal{MF}(F) \to \mathcal{ML}(F)$. For every essential simple closed curve δ the function ℓ_δ takes the same value on $(\mathcal{F}, \mu)$ as on (λ, μ'); since the ℓ_δ give the coordinates of embeddings of $\mathcal{MF}(F)$ and $\mathcal{ML}(F)$ into $\mathbb{R}^{\mathcal{S}}$, the straightening map is injective and entwines these two embeddings.

4.3.4. *Train tracks.* A geodesic lamination λ on a closed hyperbolic surface F is not dense; in fact, its support has no interior. For, if it did we could lift to the universal cover a one parameter family of leaves whose endpoints were not both constant in the family; thus at least one of the endpoints would fill out an interval $I \subset S^1_\infty$. But if we zoom in towards the midpoint of this interval by a sequence of deck transformations we could translate a fundamental domain into a region that was entirely filled by geodesics, and this would imply that the lamination was actually a nonsingular *foliation* of F, contrary to the Poincaré–Hopf formula which says this is only possible on a closed surface with $\chi = 0$. The components of $F - \lambda$ can be path completed to hyperbolic surfaces with geodesic boundary, part of which might consist of closed geodesics and part of cyclic chains of finitely many bi-infinite geodesics meeting pairwise at cusps. Typically $F - \lambda$ consists of the interior of finitely many ideal polygons, each of which is path completed by adding a single cyclic chain of bi-infinite geodesics.

Each path-completed component of $F - \lambda$ has a thick-thin decomposition, and for sufficiently small ϵ the thin part consists of rectangles foliated by intervals meeting boundary leaves of λ orthogonally. If we collapse all the thin intervals in all the complementary components the lamination collapses to a *train track* τ — a finite graph with C^1 edges whose tangent spaces match up at each vertex. The vertices of τ are the images of boundary intervals of the thin regions of $F - \lambda$. We say that τ *carries* the lamination λ. Note that some components of τ might contain no vertices; these arise from (some of the) isolated closed leaves of λ.

Suppose λ is measured by μ. If σ is a short interval transverse to an edge e of τ, the preimage of σ under this collapsing map is an efficient transversal σ' to λ. We may therefore assign a *weight* $w(e)$ to e which is the measure $\mu(\sigma')$.

At each vertex of τ if we locally orient the tangent space then some edges are incoming and some are outgoing. Invariance of the transverse measure implies that the weights satisfy the *switch conditions* at the vertex: the sum of the weights on the incoming edges equals the sum of the weights on the outgoing edges. For a given train track, the space of non-negative functions w on the edges satisfying the switch conditions is the positive cone in a finite dimensional real vector space.

We now explain how to go from a non-negative weight w on a train track τ to a measured foliation. Thicken each edge e to a Euclidean rectangle of height $w(e)$. The switch condition implies that each vertex, the sum of the heights of the incoming rectangles equals the sum of the heights of the outgoing rectangles, so we can glue the vertical sides to each other by isometries realizing this equality. The result of this gluing is a measured foliation on a subset S of F whose leaves are obtained by gluing together horizontal lines in each Euclidean rectangle, and whose invariant transverse measure is the Euclidean length of an orthogonal transversal. Each component of $F - S$ contains a spine — a graph to which this component deformation retracts in such a way that the fibers of the retraction are intervals, and so that each boundary component maps to the spine by an embedding. If we quotient each such component to a spine then S surjects onto F and its image is the desired measured foliation. Different choices of spine give rise to Whitehead equivalent measured foliations.

In this way we obtain a chain of maps

$$\mathcal{MF}(F) \xrightarrow{\text{straighten}} \mathcal{ML}(F) \xrightarrow{\text{collapse}} \text{weighted train tracks} \xrightarrow{\text{thicken}} \mathcal{MF}(F)$$

whose composition is the identity.

Remark 4.38. The collapsing operation depends on a choice of hyperbolic structure and a choice of ϵ; different train tracks carrying the same measured lamination are related by combinatorial operations that induce piecewise integral linear transformations on their spaces of weights; see e.g. Harer–Penner [10] for details. In this way spaces of train track weights induce *manifold charts* on $\mathcal{MF}(F)$ and $\mathcal{ML}(F)$ and projective weights induce manifold charts on $\mathcal{PMF}(F)$ and $\mathcal{PML}(F)$, revealing these latter spaces to be spheres of dimension $6g - 7 + 2p$, and setting the scene for the proof of Theorem 4.32.

Remark 4.39. If a lamination λ is carried by a train track τ there is a map from measures supported by λ to weights carried by τ. Because each of the chain of maps above is invertible, it follows that this map from measures to weights is also invertible. Thus the space of projective invariant measures carried by a lamination λ is a simplex whose dimension can be computed by finding a train track τ.

4.4. **Teichmüller metric.** An oriented finite area complete hyperbolic surface is ipso facto a Riemann surface, and conversely every Riemann surface R with $\chi < 0$ has a unique finite area complete hyperbolic metric in its conformal class. Thus we may define $\mathcal{T}(F)$ alternately as the space of equivalence classes of marked Riemann surfaces $f : F \to R$.

Definition 4.40 (Teichmüller metric)**.** The *Teichmüller distance* between two marked Riemann surfaces (f_1, R_1) and (f_2, R_2) is defined to be $\inf_g \log K(g)$ where $K(g)$ is the (maximal) dilatation of g, and the infimum is taken over all qc homeomorphisms $g : R_1 \to R_2$ so that gf_1 is homotopic to f_2.

The fact that this is an honest metric depends on several facts that we have already established or asserted, including:

(1) a 1-qc map is actually conformal;
(2) the composition of a K_1-qc and a K_2-qc map is K_1K_2-qc;
(3) the dilatations of g and g^{-1} are equal for any *qc*-homeomorphism g; and
(4) any two Riemann surfaces of the same genus are diffeomorphic, and any diffeomorphism between compact surfaces is quasiconformal.

Suppose R is a Riemann surface. Let $B_1(R)$ denote the space of Beltrami differentials μ on R with $\|\mu\| < 1$. This is a subset of the space of measurable sections of a normed holomorphic line bundle on R, and is therefore an open subset of a complex Banach space, and thereby may be thought of as an (infinite dimensional) complex Banach manifold. We claim there is a natural surjective map $B_1(R) \to \mathcal{T}(R)$. This may be defined as follows.

Choose an identification of the universal cover $\tilde{R}$ with $\mathbb{H}^2$. Every $\mu \in B_1(R)$ lifts to a Beltrami differential $\tilde{\mu}$ on $\mathbb{H}^2$ that may be extended by 0 to all of $\hat{\mathbb{C}}$. By Theorem 4.23 there is a quasiconformal homeomorphism $f : \hat{\mathbb{C}} \to \hat{\mathbb{C}}$ with $f_\mu = \tilde{\mu}$. The image of $\mathbb{H}^2$ is a Jordan disk, so after composition with a conformal map we may obtain $g : \mathbb{H} \to \mathbb{H}$ and satisfying $g_\mu = \tilde{\mu}$. On the one hand, such a g is unique up to composition with a conformal automorphism of $\mathbb{H}$. On the other hand, $\tilde{\mu}$ is invariant under $\pi_1(R)$, which acts by deck transformations on $\mathbb{H}$. It follows that g conjugates $\pi_1(R)$ to another discrete cocompact group Γ of conformal automorphisms of $\mathbb{H}$, and thereby determines a marked Riemann surface $R \to \mathbb{H}/\Gamma$, which is to say a point in $\mathcal{T}(R)$.

We thus obtain a surjective map $B_1(R) \to \mathcal{T}(R)$, and for $S \in \mathcal{T}(R)$ the Teichmüller distance from R to S is the infimum over $\mu \in B_1(R)$ in the fiber over S of $\log(1+\|\mu\|)/(1-\|\mu\|)$.

4.4.1. *Teichmüller maps.* It is a remarkable fact that for any two marked Riemann surfaces in the same Teichmüller space, one may write down an explicit quasiconformal map between them realizing the Teichmüller distance.

Definition 4.41 (Quadratic holomorphic differentials)**.** Let R be a Riemann surface. A *quadratic holomorphic differential* q is a holomorphic section of the (symmetric) square of the complex line bundle $T^{1,0}$. In terms of a local holomorphic coordinate z we may write q as $q(z)dz^2$.

Let q be a quadratic holomorphic differential on R. Away from the zeroes, the square root $\sqrt{q}$ is well-defined up to sign, and is a holomorphic 1-form. Integrating this 1-form gives R locally a (G, X) structure where $X = \mathbb{C}$ and G is the group of translations. The holonomy around a zero of q will be either the identity or $z \to -z$ according to whether the multiplicity of the zero is even or odd. Thus, we may pull back the foliations of $\mathbb{C}$ by horizontal and vertical lines to a pair of (singular) *measured foliations* $\mathcal{F}_s$, $\mathcal{F}_u$ on R; the measures are obtained on transversals by integrating the absolute value of the real or imaginary of $\sqrt{q}$.

Definition 4.42 (Teichmüller map)**.** Let R be a Riemann surface and let q be a quadratic holomorphic differential on R with measured foliations $\mathcal{F}_s$ and $\mathcal{F}_u$. Fix a real number $\lambda > 0$. Let S be the Riemann surface obtained from R by stretching the horizontal leaves $\mathcal{F}_u$ by a factor of λ and shrinking the vertical leaves $\mathcal{F}_s$ by a factor of λ^{-1}. The 'identity' map from R to S is a K-quasiconformal map with $K = \lambda^2$ called a *Teichmüller map.*

Theorem 4.43 (Teichmüller map realizes)**.** *Let $f : R \to S$ be a Teichmüller map. Then f is the unique quasiconformal map in its homotopy class realizing the Teichmüller distance from R to S (in the homotopy class of f).*

Proof. For concreteness suppose f is K-quasiconformal where $K = \lambda^2$ as above. The quadratic holomorphic differential ϕ defines Euclidean cone metrics on R and S which we may normalize both to have area 1. The cone angle at a zero of q of order n is $(2+n)\pi$; in particular, the Euclidean cone metric is non-positively curved, and therefore the leaves of the horizontal foliation $\mathcal{F}_u$ are *length minimizing geodesics* in the metrics both on R and on S.

Let $g : R \to S$ be another quasiconformal map homotopic to f and fix a homotopy between f and g. Since R is compact, there is a constant C so that the tracks of the homotopy have diameter at most C. Fix a positive real number $T \gg C$ and let σ be a horizontal segment of $\mathcal{F}_u$ of length T. The length of $f(\sigma)$ is λT and is a length minimizing geodesic (rel. endpoints) in S. Applying the homotopy to take $f(\sigma)$ to $g(\sigma)$, we deduce that the length of $g(\sigma)$ is at least $\lambda T - 2C$; in other words, g stretches long segments of $\mathcal{F}_u$ by a factor of at least λ.

Since there are only finitely many zeroes of q, almost every point on R is contained in a nonsingular leaf of $\mathcal{F}_u$. Thus the average length of $g(\sigma)$ over all horizontal segments of length T is equal to $T \int_R |dg|_{T\mathcal{F}_u}|\text{darea} \ge \lambda T - 2C$. Dividing by T and taking the limit as

$T \to \infty$ gives us the inequality $\int_R |dg|_{T\mathcal{F}_u}|\text{darea} \geq \lambda = K^{1/2}$. On the other hand, if g is K'-quasiconformal, the norm of the Jacobian $|Jg|$ pointwise satisfies $|Jg| \geq |dg|_{T\mathcal{F}_u}|^2/K'$. Thus, by the Cauchy–Schwarz inequality, and the fact that R and S both have area 1,

$$1 = \text{area}(S) = \int_R |Jg|\text{darea} \geq \frac{1}{K'}\int_R |dg|_{T\mathcal{F}_u}|^2\text{darea} \geq \frac{1}{K'}\left(\int_R |dg|_{T\mathcal{F}_u}|\text{darea}\right)^2 = \frac{K}{K'}$$

The inequality is strict unless $\mu_g = \mu_f$ almost everywhere, which implies $g = f$. □

Note that for a Teichmüller map f the relationship between q and μ_f is that $\mu_f = k\bar{q}/|q|$ for $k = (K-1)/(K+1)$. On a Riemann surface R of genus $g \geq 2$ the space $Q(R)$ of holomorphic quadratic differentials has complex dimension $3g-3$, which is the same dimension as Teichmüller space. For any nonzero $q \in Q(R)$ and any $1 > k \geq 0$ we may construct a unique Teichmüller map with domain R and with $\mu = k\bar{q}/|q|$. In this way we obtain a map from $\mathbb{R}^{6g-6}$ to $\mathcal{T}(R)$. This map is injective and proper by Theorem 4.43. By invariance of domain it is surjective, and therefore a homeomorphism. Thus we have proved:

Corollary 4.44. *For any two closed Riemann surfaces R and S and any homotopy class of orientation-preserving homeomorphism between them, there is a (unique) Teichmüller map $f : R \to S$ in the given homotopy class.*

A similar statement holds for surfaces with punctures.

4.4.2. *Length and modulus.* Suppose (f_1, S_1) and (f_2, S_2) are marked hyperbolic (and therefore also Riemann) surfaces in $\mathcal{T}(F)$. How may we estimate their Teichmüller distance? The following Lemma is very useful:

Lemma 4.45 (Length and modulus)**.** *Let S be a complete hyperbolic surface and let γ be a closed geodesic in S. Let S_γ be the (annular) covering space of S associated to the $\mathbb{Z}$ subgroup of $\pi_1(S)$ generated by the conjugacy class of γ. Then*

$$\text{length}(\gamma) = \pi/m(S_\gamma)$$

Proof. In the upper half-space model a fundamental domain for S_γ is bounded by two semicircles centered at 0, one of radius 1 and one of radius $e^{\text{length}(\gamma)}$. Taking logs maps this region to a rectangle with side lengths π and length(γ). Gluing the two opposite sides with length π makes an annulus conformally isomorphic to S_γ; thus $m(S_\gamma) = \pi/\text{length}(\gamma)$. □

Here is a corollary.

Corollary 4.46 (Length and dilatation)**.** *Suppose we have two points (f_1, S_1) and (f_2, S_2) in $\mathcal{T}(F)$. For any essential closed curve δ in F let γ_i be the geodesic representative of $f_i(\delta)$ in S_i and by abuse of notation let $\ell_\delta(S_i)$ be the respective lengths of the γ_i. If the Teichmüller distance from (f_1, S_1) to (f_2, S_2) is K, then*

$$\ell_\delta(S_2)/K \leq \ell_\delta(S_1) \leq K\ell_\delta(S_2)$$

Proof. Lift a K-qc map $g : S_1 \to S_2$ to a map between their respective annular covering spaces. Since the property of being K-qc is local, this lifted map is K-qc and therefore distorts the modulus of annuli by at most a factor of K. □

From this one may see that the topology on $\mathcal{T}(F)$ induced by length functions and the topology induced by the Teichmüller metric agree.

4.4.3. *Action of the mapping class group.* The mapping class group $\mathrm{Mod}(F)$ acts on $\mathcal{T}(F)$; if $\varphi \in \mathrm{Mod}(F)$ and (f, S) is a marked hyperbolic surface, then $(f \circ \varphi^{-1}, S)$ is a marked hyperbolic surface, where by abuse of notation we let $f \circ \varphi^{-1}$ denote the homotopy class of any map obtained by composing f with a representative of φ^{-1}. This action is discrete and properly discontinuous though not free since there are hyperbolic surfaces which admit a nontrivial (finite) group of isometries.

The quotient $\mathcal{M}(F) := \mathcal{T}(F)/\mathrm{Mod}(F)$ is the moduli space of (unmarked) hyperbolic structures on F. It is an orbifold, with orbifold fundamental group $\pi_1^o(\mathcal{M}(F)) = \mathrm{Mod}(F)$.

5. Kleinian groups

5.1. **Kleinian groups.**

Definition 5.1 (Kleinian group)**.** A *Kleinian group* Γ is a discrete, finitely generated subgroup of isometries of $\mathbb{H}^3$

Associated to a Kleinian group Γ there is a natural decomposition of S^2_∞ into two subsets canonically associated to Γ; the (closed) *limit set* $\Lambda(\Gamma)$, and the (open) *domain of discontinuity* $\Omega(\Gamma) := S^2_\infty - \Lambda(\Gamma)$. We denote these by Λ and Ω if Γ is understood.

In the Poincaré ball model, the union $\mathbb{H}^3 \cup S^2_\infty$ is a closed unit ball. For any $x \in \mathbb{H}^3$ we can form the orbit Γx and take the closure $\overline{\Gamma x}$ in the closed unit ball. The limit set is then the difference $\Lambda(\Gamma) = \overline{\Gamma x} - \Gamma x$ which is equal to $\overline{\Gamma x} \cap S^2_\infty$ because Γx is discrete in $\mathbb{H}^3$. Note that this set does not depend on the choice of point x, since if $g_i x \to p \in S^2_\infty$ then $g_i y \to p$ for any other $y \in \mathbb{H}^3$.

A Kleinian group is said to be *elementary* if the limit set contains at most 2 points. This is equivalent to the group being virtually abelian (i.e. containing an abelian subgroup of finite index).

Lemma 5.2. *If Γ is not elementary, Λ is the unique minimal closed non-empty Γ-invariant subset of S^2_∞.*

Proof. Associated to any closed invariant subset $K \subset \mathbb{H}^3 \cup S^2_\infty$ we can form the convex hull $C(K)$ which is the intersection of all closed half spaces that contain K. Since Γ is non-elementary, K contains more than one point, so $C(K)$ contains some point x in $\mathbb{H}^3$. Since K is invariant, so is $C(K)$, and therefore $C(K)$ contains Γx and (since it is closed) $\overline{\Gamma x}$ and therefore Λ. □

Corollary 5.3. *If Γ' is a normal subgroup of Γ (both nonelementary), then $\Lambda(\Gamma') = \Lambda(\Gamma)$.*

Proof. Since Γ conjugates Γ' to itself, it takes $\Lambda(\Gamma')$ to itself. Since Γ' is nonelementary, $\Lambda(\Gamma')$ is nonempty. □

Example 5.4. If Γ has finite covolume then $\Lambda = S^2_\infty$. This is the case for Γ the deck group of the universal cover of any complete finite volume hyperbolic 3-manifold (for example, any closed hyperbolic 3-manifold).

Example 5.5. Suppose M is a closed or finite-volume complete hyperbolic 3-manifold fibering over the circle (one example is the figure 8 knot complement). If $S \to M$ denotes inclusion of the fiber, there is a short exact sequence of fundamental groups

$$0 \to \pi_1(S) \to \pi_1(M) \to \mathbb{Z} \to 0$$

so that the fundamental group of the fiber is normal in $\pi_1(M)$. Since M is finite volume, $\Lambda(\pi_1(M)) = S^2_\infty$ and therefore also $\Lambda(\pi_1(S)) = S^2_\infty$. This is despite the fact that $\mathbb{H}^3/\pi_1(S)$ is an infinite cyclic cover of M, and therefore has infinite volume.

5.1.1. *Torsion.* According to the definition, a Kleinian group Γ may have torsion. The quotient $\mathbb{H}^3/\Gamma$ is therefore in general a complete hyperbolic 3-orbifold. It turns out that every such orbifold is good — i.e. it has a finite manifold cover. Equivalently, every Kleinian group Γ contains a finite index subgroup which is torsion-free. This fact is usually called *Selberg's Lemma*; we state and prove this in a slightly more general form than is usual.

Theorem 5.6 (Selberg's Lemma)**.** *Let Γ be a finitely generated subgroup of $\mathrm{SL}(2, \mathbb{C})$ and let $t \neq 2$. Then there exists a finite index normal subgroup N of Γ such that no element of N has trace t.*

Proof. Let R be the ring generated by the matrix entries of the generators of Γ and their inverses (we assume without loss of generality that $t \in R$, or the lemma is vacuous). Now, it is a fact that for any subring R of $\mathbb{C}$ finitely generated over $\mathbb{Z}$, that the intersection of the maximal ideals of R is equal to 0, and for any maximal ideal $\mathfrak{m}$ the quotient $R/\mathfrak{m}$ is a finite field. Since $t \neq 2$ let $\mathfrak{m}$ be a maximal ideal which does not contain $t - 2$, and define $K := R/\mathfrak{m}$. Then there is a natural map $\phi : \mathrm{SL}(2, R) \to SL(2, K)$ obtained by reducing entries mod $\mathfrak{m}$, and the kernel intersects Γ in a finite index subgroup N.

Every element of N can be written in the form $I + A$ where A is a matrix with entries in $\mathfrak{m}$. If such a matrix had trace t then $\mathrm{tr}(A) = t - 2$. But A has entries in $\mathfrak{m}$, so $\mathrm{tr}(A)$ is in $\mathfrak{m}$, contrary to the hypothesis that $t - 2$ is not in $\mathfrak{m}$. Thus N is the desired subgroup. □

Corollary 5.7. *A finitely generated Kleinian group has only finitely many conjugacy classes of torsion elements, and contains a torsion-free subgroup of finite index.*

Proof. Without loss of generality we may assume Γ is orientation-preserving. A torsion element of order n is elliptic, and is a power of an element with trace $2\cos(\pi/n)$. A ring R containing $2\cos(\pi/n_j)$ for infinitely many integers n_j is infinitely generated; now apply Theorem 5.6. □

In the sequel for simplicity we shall usually restrict our attention to Kleinian groups which are torsion-free and orientation-preserving. For such a group, the quotient $\mathbb{H}^3/\Gamma$ is a complete oriented hyperbolic 3-manifold (typically of infinite volume).

5.1.2. *Quotient manifolds.* Let Γ be a torsion-free non-elementary Kleinian group and let $C(\Lambda)$ denote the convex hull of the limit set $\Lambda(\Gamma)$. Since $C(\Lambda)$ is convex, and distance to any point $p \in \mathbb{H}^3 - C(\Lambda)$ is a convex function, there is a unique point $r(p) \in \partial C(\Lambda)$ which is closest to p. The function r varies continuously on $\mathbb{H}^3$ and extends continuously in a unique way to $\Omega(\Gamma)$: for $p \in \Omega(\Gamma)$ the point $r(p)$ is the unique point contained in a horosphere centered at p that intersects $C(\Lambda)$ only at $r(p)$. The map $r : \mathbb{H}^3 \cup \Omega(\Gamma) \to C(\Lambda)$

is proper, distance non-increasing (in $\mathbb{H}^3$) and Γ equivariant. It is called the *nearest point map*.

Since the action of Γ on $\mathbb{H}^3$ (and therefore also on $C(\Lambda)$ is free and properly discontinuous, the same is true of the action on $\mathbb{H}^3 \cup \Omega(\Gamma)$. This suggests taking three natural quotients:

(1) $C(\Gamma) := C(\Lambda)/\Gamma$, the *convex core*;
(2) $M(\Gamma) := \mathbb{H}^3/\Gamma$, the *complete hyperbolic manifold*; and
(3) $N(\Gamma) := (\mathbb{H}^3 \cup \Omega(\Gamma))/\Gamma$, the *Kleinian manifold.*

There are inclusions $C(\Gamma) \subset M(\Gamma) \subset N(\Gamma)$, all of which are homotopy equivalences. Each of $M(\Gamma)$ and $N(\Gamma)$ is a 3-manifold with fundamental group Γ. The manifold $M(\Gamma)$ is homeomorphic to the interior of $N(\Gamma)$ which may or may not be closed, and has (possibly empty) boundary $\Omega(\Gamma)/\Gamma$. Notice that a boundary component of $N(\Gamma)$ is incompressible in $N(\Gamma)$ if and only if it is covered by a simply connected component of Ω. The nearest point map r descends to homotopy equivalent deformation retractions $r : M(\Gamma) \to C(\Gamma)$ and $r : N(\Gamma) \to C(\Gamma)$.

Example 5.8. The quotient $C(\Gamma)$ is not always a manifold. For example, let F be a complete finite area hyperbolic surface, and let $\Gamma \subset \mathrm{PSL}(2,\mathbb{R})$ be a discrete group with $\mathbb{H}^2/\Gamma = F$. Then Γ is a Kleinian subgroup of $\mathrm{PSL}(2,\mathbb{C})$, and Λ is a round circle stabilized by Γ. Thus $C(\Gamma)$ is a totally geodesic copy of $\mathbb{H}^2$ in $\mathbb{H}^3$ and $C(\Gamma)$ is a totally geodesic surface isometric to F in the complete hyperbolic manifold $M(\Gamma)$ which is itself homeomorphic to $F \times \mathbb{R}$.

5.1.3. *Geometrically finite groups.*

Definition 5.9 (Geometrically finite)**.** A Kleinian group Γ is *geometrically finite* if the ϵ-neighborhood of $C(\Gamma)$ in $M(\Gamma)$ has finite volume.

If Γ does not contain parabolic elements, geometrically finite is the same as *convex cocompact* — i.e. that $C(\Gamma)$ is compact. Whether there are parabolics or not, geometrically finite is equivalent to compactness of the thick part of $C(\Gamma)$.

If Γ does not contain parabolic elements, it is geometrically finite if and only if the Kleinian manifold $N(\Gamma)$ is compact, because the nearest point map $r : N(\Gamma) \to C(\Gamma)$ is proper so one is compact if and only if the other is.

Suppose Γ is convex cocompact. In this case, the neighborhood $N_\epsilon(C(\Gamma))$ is *uniformly strictly* convex (i.e. there are uniform positive lower bounds on the principle curvatures into the neighborhood). This property is stable under perturbation; i.e. for nearby structures, the image of $N_\epsilon(C(\Gamma))$ will still be uniformly strictly convex, and will therefore contain the convex hull for the deformed structure (which must, *a posteriori*, be compact). It follows that the property of being convex cocompact is stable under perturbation. Many interesting examples of geometrically finite Kleinian groups can be obtained by deformation as we shall shortly see.

5.1.4. *Spaces of representations.* Let K be a finite CW complex which is a $K(\pi,1)$; the examples we care about will be compact irreducible atoroidal 3-manifolds, and surfaces with $\chi < 0$. Let $H(K)$ denote the set of isometry classes of complete hyperbolic 3-manifolds marked by K; i.e. points in $H(K)$ are equivalence classes of pairs (f,M) where $f : K \to M$ is a homotopy equivalence, and where $(f_1,M_1) \sim (f_2,M_2)$ if there is an isometry $\phi : M_1 \to M_2$ such that ϕf_1 is homotopic to f_2.

An equivalent definition of $H(K)$ is that it is the set of conjugacy classes of discrete faithful representations from $\pi_1(K)$ to the group of isometries of $\mathbb{H}^3$. In the sequel we usually restrict attention to oriented 3-manifolds and orientation-preserving isometries without comment. The *algebraic topology* on $H(K)$ is the topology of convergence (up to conjugacy) on generators of $\pi_1(K)$. Let $AH(K)$ denote the topological space whose underlying set is $H(K)$ with the algebraic topology.

Let us suppose that $\pi_1(K)$ is not virtually abelian, so that $\Gamma := \rho(\pi_1(K))$ is not elementary for any $\rho \in AH(K)$. If we fix a finite generating set $g_1, \cdots, g_n$ for $\pi_1(K)$ then ρ determines a canonical basepoint in $\mathbb{H}^3$ as follows. Let $\gamma_j := \rho(g_j)$ for each j. For each hyperbolic element $\gamma \in \mathrm{PSL}(2,\mathbb{C})$ the function $d(p,\gamma(p))$ is convex, and is minimized only on the axis of γ; the story for parabolic elements is similar: $d(p,\gamma(p))$ is strictly convex and converges to the minimum value (i.e. zero) at the unique fixed point at infinity. Since Γ is nonelementary, some distinct generators have distinct axes or fixed points at infinity if they are parabolic. Thus $\sum_j d(p,\gamma_j(p))$ is strictly convex and proper and is minimized at a unique basepoint $p \in \mathbb{H}^3$.

One corollary is that $AH(K)$ is Hausdorff:

Proposition 5.10 (Hausdorff)**.** *The space $AH(K)$ is Hausdorff.*

Proof. If ρ_i is a sequence of representations converging to ρ, with conjugates $\rho_i^{\alpha_i}$ converging to ρ', then the basepoints p_i for ρ_i converge to the basepoint p for ρ and the basepoints q_i for $\rho_i^{\alpha_i}$ converge to the basepoint q for ρ'.

But then the conjugating elements α_i take p_i to q_i and therefore stay in a compact subset of $\mathrm{PSL}(2,\mathbb{C})$, so some subsequence converges and conjugates ρ to ρ'. □

Another way to see that $AH(K)$ is Hausdorff is to think of $AH(K)$ as a subset of $\mathrm{Hom}(\pi_1(K),\mathrm{PSL}(2,\mathbb{C}))$/conjugacy, where the latter is given the quotient topology. This topology is not Hausdorff, but is Hausdorff away from the reducible representations, but a non-elementary Kleinian group is never reducible.

It is convenient to observe that $AH(K)$ is closed as a subset of the space of representations up to conjugacy. This is known as Chuckrow's Theorem:

Proposition 5.11 (Chuckrow's Theorem)**.** *If $\pi_1(K)$ is not abelian then $AH(K)$ is closed in* $\mathrm{Hom}(\pi_1(K),\mathrm{PSL}(2,\mathbb{C}))$*/conjugacy.*

Proof. Another way to say this is that if $\rho_i \to \rho$ and if ρ_i are discrete and faithful then so is ρ. Suppose not; then whether ρ is indiscrete or unfaithful or both there is a sequence of nontrivial elements $g_i \in \pi_1(K)$ where $\rho_i(g_i) \to \mathrm{id}$. Then for any $h \in \pi_1(K)$ we have $\rho_i(h) \to \rho(h)$ and therefore $\rho_i([g_i,h]) \to \mathrm{id}$. Thus by Margulis' Lemma and the fact that $\pi_1(K)$ is torsion free, g_i and $[g_i,h]$ commute for sufficiently large i.

But for non-elliptic elements of $\mathrm{PSL}(2,\mathbb{C})$ if g_i and $[g_i,h]$ commute then g_i and h commute; thus for any two $h,h' \in \pi_1(K)$ there is an i so that both h and h' commute with g_i. Since each ρ_i is discrete and faithful this implies that h and h' commute with each other, but since they were arbitrary this implies $\pi_1(K)$ is abelian after all. □

5.2. **Examples of Kleinian groups.** Let us now consider some examples of Kleinian groups. The simplest examples are *Schottky groups*.

5.2.1. *Schottky groups.* Let $\gamma_1, \cdots, \gamma_n$ be hyperbolic isometries with repelling/attracting fixed points $p_1^\pm, \cdots, p_n^\pm$.

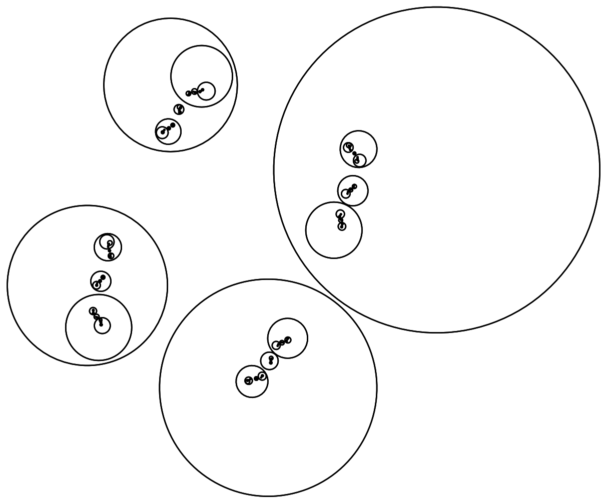

FIGURE 13. Dividing circles whose translates under the elements of a Schottky group accumulate on a Cantor limit set.

Suppose there are disjoint topological disks $D_j^\pm \subset S^2_\infty$ where $D_j^\pm$ contains $p_j^\pm$ in the interior but no other $p_k^\pm$, and suppose that γ_j takes ∂D_j^- to ∂D_j^+ for each j. Let $P := S^2 - \cup D_j^\pm$ and let Γ be the group generated by the γ_j. Then each γ_j^+ takes $S^2 - D_j^-$ entirely inside D_j^+ and similarly for γ_j^-. It follows that Γ is freely generated by the γ_j, and P is a fundamental domain for the action of Γ on $\Omega = \Gamma \cdot P$. The limit set Λ is a Cantor set, and the quotient $N(\Gamma)$ is a genus g handlebody with boundary Ω/Γ a genus g surface. See Figure 5.2.1.

5.2.2. *Quasifuchsian groups.* Suppose F is a closed oriented surface, and $\rho : \pi_1(F) \to \mathrm{PSL}(2, \mathbb{C})$ is discrete and faithful. Let $\Gamma := \rho(\pi_1(F))$ and let $M(\Gamma) = \mathbb{H}^3/\Gamma$. The representation ρ is *Fuchsian* if the limit set Λ is a round circle, and *quasifuchsian* if Λ is a quasicircle. The domain of discontinuity Ω consists of two topological disks $\Omega^\pm$, each of which is separately stabilized by Γ. Thus the Kleinian manifold $N(\Gamma)$ has two boundary components $R^\pm := \Omega^\pm/\Gamma$, each homotopy equivalent and therefore also homeomorphic to F, and the inclusion of each of these boundary components into $N(\Gamma)$ is a homotopy equivalence. In particular, there is a proper map $F \times [0, 1] \to N(\Gamma)$ with degree 1 on the boundary components. Thus $N(\Gamma)$ is compact (and in fact homeomorphic to $F \times [0, 1]$ since it is Haken) and Γ is geometrically finite.

The space of quasifuchsian representations up to conjugacy is denoted $\mathfrak{QF}(F)$. The surfaces R^+ and R^- are Riemann surfaces, which come with a marking by F and $\bar{F}$ (i.e. F with the opposite orientation) respectively. Thus each quasifuchsian representation gives rise to a point $(R^+, R^-) \in \mathfrak{T}(F) \times \mathfrak{T}(\bar{F})$ (our notation omits the marking for simplicity).

Theorem 5.12 (Bers simultaneous uniformization)**.** *The map $\mathfrak{QF}(F) \to \mathfrak{T}(F) \times \mathfrak{T}(\bar{F})$ is a homeomorphism.*

Proof. We obtain an inverse via the measurable Riemann mapping theorem. Let $(S^+, S^-) \in \mathcal{T}(F) \times \mathcal{T}(\bar{F})$ be arbitrary (we suppress markings for simplicity) and let μ^+, μ^- be Beltrami differentials on R^+ and R^- associated to quasiconformal maps $R^\pm \to S^\pm$. Lift $\mu^\pm$ to Beltrami differentials on $\Omega^\pm$ to get a single Beltrami differential μ on S^2_∞, and let $f_\mu : S^2_\infty \to S^2_\infty$ be the associated measurable Riemann map. Then f_μ conjugates Γ to a quasifuchsian group with parameter (S^+, S^-). □

Example 5.13 (Bers slice). Fix $Y \in \mathcal{T}(\bar{F})$. The subspace $B_Y \subset \mathcal{QF}(F)$ consisting of representations for which the projection to the second factor $\mathcal{QF}(F) \to \mathcal{T}(F) \times \mathcal{T}(\bar{F}) \to \mathcal{T}(\bar{F})$ maps to Y, is (evidently) homeomorphic to $\mathcal{T}(F)$. This embedding of $\mathcal{T}(F)$ into $\mathcal{QF}(F)$ is called a *Bers slice.*

Let $\Gamma \in \mathcal{QF}(F)$. Associated to each essential simple closed curve δ in F there are three numbers:

(1) the length of the geodesic representative of δ in the hyperbolic structures on $\mathbb{R}^\pm$ (denote these by $\ell^\pm_\delta$); and
(2) the length of the geodesic representative of δ in the hyperbolic 3-manifold M (denote this by $\ell_\delta(M)$).

These numbers are related as follows:

Lemma 5.14 (Bers inequality)**.** *For any quasifuchsian representation Γ of $\pi_1(F)$ and any essential simple closed curve δ on F there is an inequality*

$$\frac{1}{\ell^+_\delta} + \frac{1}{\ell^-_\delta} \le \frac{2}{\ell_\delta(M)}$$

In particular, $\ell_\delta(M) \le 2\min(\ell^+_\delta, \ell^-_\delta)$.

Proof. Let $\gamma \in \Gamma$ be the image of δ. This element acts freely on $S^2_\infty - \mathrm{fix}(\gamma)$ and a fundamental domain is a round annulus A of modulus $\ell_\delta(M)/2\pi$ by Lemma 4.45. The quotient by γ is a torus T, which is divided by the image of Λ (which consists of two longitudinal curves) into two annuli $A^\pm$ whose moduli are π/ℓ^+_δ and π/ℓ^-_δ respectively, again by Lemma 4.45. On the other hand, as in Example 4.12 extremal length gives the inequality $m(A^+) + m(A^-) \le 1/m(A)$. The lemma follows. □

Corollary 5.15. *The closure of a Bers slice $B_Y(F)$ in $AH(F)$ is compact.*

Proof. By Bers inequality the trace of each element of $\pi_1(F)$ is uniformly bounded throughout $B_Y(F)$. Therefore the closure of $B_Y(F)$ is compact in the space of representations up to conjugacy. □

Example 5.16 (Punctured torus slice). The discussion above goes through for surfaces F with punctures if we restrict attention to representations for which the cusps map to parabolic elements. Let's consider the case of a once-punctured torus.

Just as for closed surfaces, one has the Bers isomorphism $\mathcal{QF}(F) \to \mathcal{T}(F) \times \mathcal{T}(\bar{F})$. In the case that F is a punctured torus, $\mathcal{QF}(F)$ is 2 complex dimensional. Let us describe a particular one complex dimensional slice (not a Bers slice!!)

We fix free meridian and longitude generators a, b for $\pi_1(F)$. For a discrete representation in which both elements are loxodromic, they each have two distinct fixed points, and we shall cut out a 1 complex dimensional slice of $\mathcal{QF}(F)$ by taking the fixed points of a to be ± 1 and the fixed points of b to be $\pm i$. This condition on fixed points selects representations of the form

$$\rho(a) = \frac{1}{\sqrt{\zeta^2 - 1}} \begin{pmatrix} \zeta & 1 \\ 1 & \zeta \end{pmatrix} \quad \rho(b) = \frac{1}{\sqrt{\omega^2 - 1}} \begin{pmatrix} \omega & i \\ -i & \omega \end{pmatrix}$$

while the condition that $\rho([a,b])$ is parabolic imposes the relation $\omega = -\zeta/\sqrt{\zeta^2 - 1}$. Let $\Gamma(\zeta)$ be the group generated by these elements, and let $B \subset \mathbb{C}$ denote the space of parameters ζ for which $\Gamma(\zeta)$ is quasifuchsian. Notice that the closure of B in the space of representations is compact away from $\zeta \to 1$ and $\zeta \to \infty$ where the traces of $\rho(a)$ and $\rho(b)$ respectively go to infinity.

The map $z \to 1/z$ on $\mathbb{CP}^1$ conjugates $\Gamma(\zeta)$ to itself and acts on $H_1(F)$ by the (orientation reversing!) element $\left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$ i.e. it fixes $\rho(a)$ and takes $\rho(b)$ to $\rho(b^{-1})$ (similarly the map $z \to -1/z$ fixes $\rho(b)$ and sends $\rho(a)$ to $\rho(a^{-1})$). Consequently it interchanges the two sides of the limit set, and shows that (in Bers coordinates on $\mathcal{QF}(F)$) the slice B is the graph of the map $\iota : \mathcal{T}(F) \to \mathcal{T}(\bar{F})$ that is the composition of complex conjugation with the elliptic involution which acts on $H_1(F)$ by multiplication by -1. However, unlike the Fuchsian slice, which is the graph of complex conjugation alone and which is a totally real subset of $\mathcal{QF}(F)$, the slice B is complex.

If ζ is real and has absolute value > 1 this representation is Fuchsian, and stabilizes the unit circle in $\mathbb{C}$. When we leave the real axis by adding a small imaginary part to ζ, the limit set deforms to a quasicircle. See Figure 14.

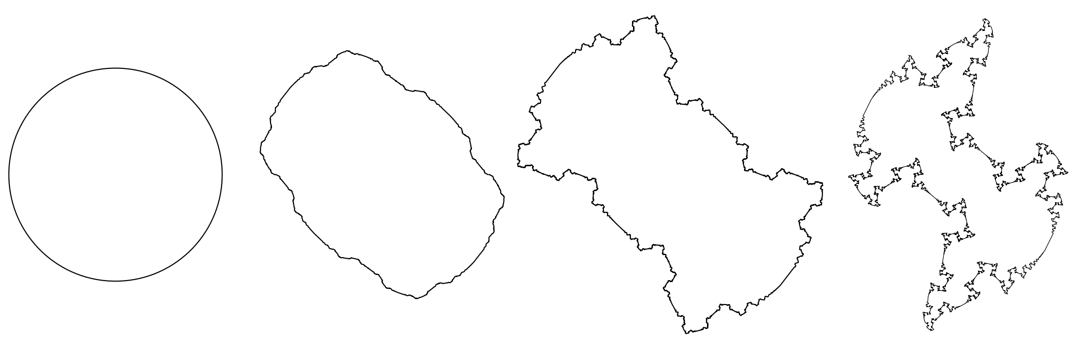

FIGURE 14. Quasicircular limit sets associated to the parameters $\zeta = 1.3$, $\zeta = 1.3 + 0.2i$, $\zeta = 1.15 + 0.4i$ and $\zeta = 0.9 + 0.4i$.

For each essential simple closed curve on F represented by a conjugacy class $w \in \pi_1(F)$ the trace $\mathrm{tr}(w)$ is real on the interval $\zeta \in (1, \infty)$ and its absolute value is convex with a unique minimum which is a critical point of $\mathrm{tr}'(w)$. The absolute value of $\mathrm{tr}(w)$ has a saddle at this critical point, and the descending gradient flowlines may be characterized by the property that $\mathrm{tr}(w)$ is real. Follow this descending flowline until $|\mathrm{tr}(w)| = 2$ at a point on the frontier of B.

At this limit point the group develops an *accidental parabolic*, and the limit quasicircle is pinched on one side along the translates of the lifts of the simple closed curve into a collection of infinitely many mutually tangent circles, and the associated Kleinian group is called a *cusp group*. Because of the symmetry, the limit quasicircle is simultaneously

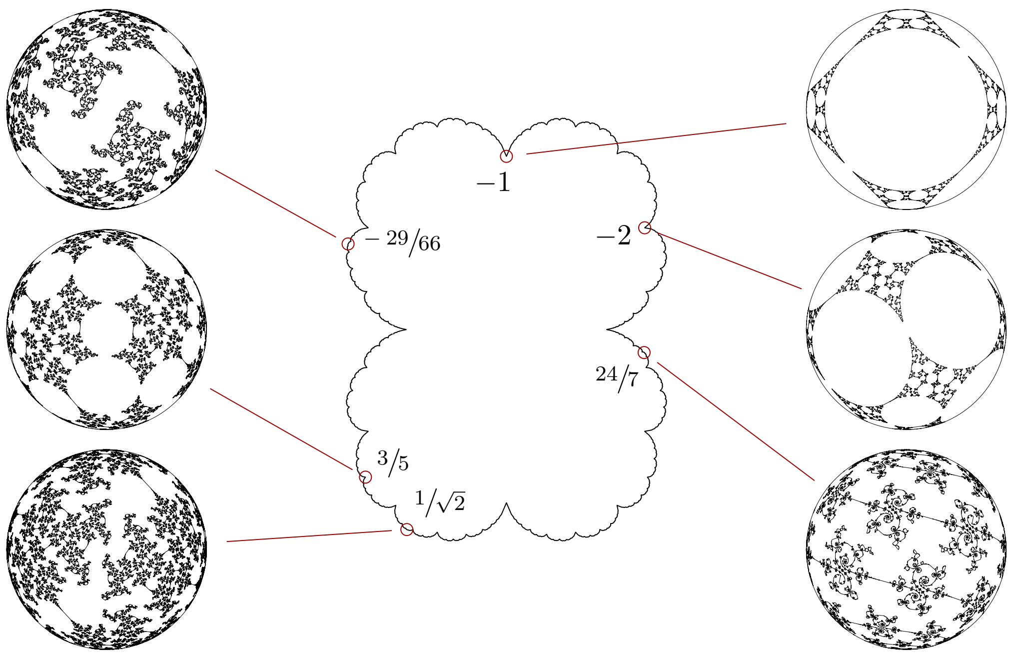

FIGURE 15. Quasicircles degenerate to infinite circle packings as essential simple curves pinch and the representations develop 'accidental' parabolics. When the degeneration occurs along an irrational lamination the limit is more complicated; if the lamination fills the surface (which holds automatically for a once-punctured torus) the limit is totally degenerate. In this family because of symmetry the 'same' degeneration occurs on both sides.

pinched along the *other* side along the translates of the lifts of the simple closed curve $\bar{w}$ which is obtained from w by replacing b with b^{-1}. These groups are still geometrically finite, but not quasifuchsian.

If w_i is a sequence of essential simple closed curves with slopes approaching an irrational number θ, the sequence of associated cusp groups converges to a degenerate limit. Because of symmetry, if the limit set degenerates one side it degenerates on both, so the only possibility for an irrational limiting representation is that the limit set converges to a sphere filling curve.

For example, taking $\theta = 1/\sqrt{2}$ the parameters converge to $\zeta = e^{i\pi/4}$ which is a fiber of the once punctured torus bundle with monodromy $\left(\begin{smallmatrix} 3 & 4 \\ 2 & 3 \end{smallmatrix}\right)$.

Figure 15 depicts the frontier of this slice of $\mathcal{QF}(F)$, transformed to a bounded domain by $\zeta \to 1/\zeta^2 - 1/2$. The boundary is a Jordan curve, continuously parameterized by $\mathbb{RP}^1$ which may be identified with the space $\mathcal{PMF}(F)$ of slopes on F. Several specific points on the boundary are labeled, together with the limit set (on the sphere) of the corresponding group.

5.3. **Dynamics at infinity.**

5.3.1. *Harmonic functions.* A function on a Riemannian manifold is *harmonic* if its gradient flow is volume preserving; equivalently if its average value on infinitesimal round balls is equal to the value at the center. On a hyperbolic manifold this mean value property can be integrated: a function on $\mathbb{H}^n$ is harmonic if its average value on *every* round ball is equal to the value at the center. Suppose h is harmonic on $\mathbb{H}^n$ and limits measurably to a non-negative bounded measurable function f on the boundary. Then the mean value property says that the value at every point is the visual average of the boundary values; i.e. if we identify S^{n-1}_∞ with the unit tangent sphere UT_p by exponentiating, $h(p)$ should be equal to the average value of f on S^{n-1}_∞ with respect to the (exponentiated push forward of the) Riemannian measure on UT_p.

In fact the converse is also true: given any non-negative bounded measurable function f on S^{n-1}_∞ we may obtain a harmonic function h_f with these boundary values by visual averaging. We claim that the function h_f so obtained is harmonic. Since linear combinations of harmonic functions are harmonic, it suffices to prove this for a function f on S^{n-1}_∞ equal to $\epsilon^{-(n-1)}$ on a round ball of radius ϵ and 0 elsewhere. The visual measure of this ball is very nearly constant on horospheres centered at the center of the ball, and decays with a factor of $e^{-(n-1)d}$ with distance d. If we put this ball at infinity in the upper half space model and take a limit, the gradient flows of the visual extensions converge to the vector field $z^{n-1}\partial/\partial z$ (where z is the vertical coordinate) which is volume preserving.

One important use of harmonic functions is to compare dynamical information in the quotient manifold and at infinity.

Theorem 5.17 (Ahlfors)**.** *If Γ is geometrically finite then Λ has either zero measure or full measure. If Λ has full measure, the action of Γ on S^2_∞ is ergodic.*

To say that Γ acts ergodically on S^2_∞ is to say that every Γ-invariant measurable set has either zero or full Lebesgue measure (implicit in this is the fact that the action of Γ preserves the measure class of Lebesgue measure).

Proof. It is equivalent to show that any bounded measurable function f supported on Λ and invariant by Γ is constant a.e. with respect to Lebesgue measure. Let f be such a function, and let h_f be the harmonic extension to $\mathbb{H}^3$. Without loss of generality we may assume f takes only the values 0 and 1. Now, if $x \in \mathbb{H}^3$ is outside $C(\Lambda)$ then $h_f(x) < 1/2$, since there is a hemisphere containing x and missing Λ. Since f is Γ-invariant, so is h_f, so it descends to a function (which by abuse of notation we also call h_f) on M. The subset where $h_f \geq 1/2$ is contained in the convex core $C(\Gamma)$ which has finite volume. On the other hand h_f is harmonic, and therefore its gradient flow is volume preserving. But a gradient flow takes the subset where $h_f \geq 1/2$ inside itself and takes it strictly inside itself (which would contradict finite volume) if there is any point where $h_f = 1/2$. Therefore $h_f < 1/2$ everywhere. But this means h_f and therefore f is identically zero, since near a point of density for the support of f we would have h_f close to 1. □

5.3.2. *Geometric convergence.* A sequence of Kleinian groups Γ_i is said to converge *geometrically* if they converge in the Hausdorff topology in the space of closed subsets of $\mathrm{PSL}(2,\mathbb{C})$. The limit Γ_∞ is necessarily a subgroup. It might not be discrete; however by the Margulis Lemma the only possible indiscrete limits are virtually abelian. And it might not be finitely generated; e.g. a sequence of finite index subgroups of a fixed Γ converging to an infinite index subgroup. Notice that here we really mean to think of the Kleinian groups as honest subgroups of a fixed Lie group, and not just as conjugacy classes.

If we fix an orthogonormal frame $\tilde{e}$ at a basepoint $\tilde{p}$ in $\mathbb{H}^3$ the image in the quotient manifold $M_i = \mathbb{H}^3/\Gamma_i$ is an orthonormal frame e_i at some basepoint p_i in M_i. Conversely, the data of Γ_i can be recovered from the pair (M_i, e_i). Geometric convergence of Γ_i is convergence in the Gromov–Hausdorff topology of the sequence (M_i, e_i). Note that if the injectivity radius at the basepoint is uniformly bounded below (for instance if the e_i all lie in the thick part of M_i), any sequence has a convergent subsequence.

Lemma 5.18. *Let $\Gamma_i \to \Gamma_\infty$ be a geometrically convergent sequence of Kleinian groups, where the limit Γ_∞ is discrete (but not necessarily finitely generated), and let $M_i := \mathbb{H}^3/\Gamma_i$ and $M_\infty := \mathbb{H}^3/\Gamma_\infty$. Denote the limit sets of Γ_i by Λ_i and the limit set of Γ_∞ by Λ_∞.*

1. *there is an inclusion $\Lambda_\infty \subset \liminf \Lambda_i$;*
2. *if the injectivity radius in the convex cores $C(\Gamma_i) := C(\Lambda_i)/\Gamma_i$ is uniformly bounded above by some constant R then the same is true for Γ_∞, and $\Lambda_\infty = \lim \Lambda_i$.*

Proof. Since limit sets are minimal, the set of fixed points of hyperbolic elements are always dense in any limit set. These points are stable under perturbation. Thus $\Lambda_\infty \subset \liminf \Lambda_i$.

Now suppose the injectivity radius is the convex cores $C(\Gamma_i)$ are uniformly boundary above by some constant R. For any complete hyperbolic manifold $M = \mathbb{H}^3/\Gamma$ let M_R be the subset where the injectivity radius is bounded above by R, and let $\tilde{M}_R \subset \mathbb{H}^3$ be its universal cover. We claim that the closure of $\tilde{M}_R$ in S^2_∞ is contained in $\Lambda(\Gamma)$. For, every point in $\tilde{M}_R$ is either a uniformly bounded distance from a closed geodesic of definite length (and therefore it is within bounded distance of the convex hull) or it is within uniformly bounded distance from the thin part, which itself accumulates at infinity only at (a subset of) Λ.

Evidently $\limsup C(\Gamma_i) \subset \limsup (M_i)_R \subset (M_\infty)_R$ since injectivity radius cannot jump up at a limit point. It follows that $\limsup \Lambda_i \subset \Lambda_\infty$ so $\Lambda_\infty = \lim \Lambda_i$. Consequently $C(\Gamma_\infty) = \limsup C(\Gamma_i) \subset (M_\infty)_R$ so the injectivity radius in $C(\Gamma_\infty)$ is bounded above by R and the proposition is proved. ∎

5.3.3. *No invariant line fields.* Suppose $f : M \to M'$ is a quasi-isometric homeomorphism between complete hyperbolic 3-manifolds $M = \mathbb{H}^3/\Gamma$ and $M' = \mathbb{H}^3/\Gamma'$. Recall, as in the proof of Mostow Rigidity, that the map f lifts to $\tilde{f} : \mathbb{H}^3 \to \mathbb{H}^3$ and extends to a quasiconformal homeomorphism $\tilde{f}_\infty : S^2_\infty \to S^2_\infty$ conjugating Γ to Γ'. Conversely, any quasiconformal conjugacy between Γ and Γ' extends to $\mathbb{H}^3$ and covers a quasi-isometric homeomorphism from M to M'. See Theorem 4.24 and the following remark.

Quasiconformal conjugacies supported in $\Omega(\Gamma)$ are parameterized by the Teichmüller space of Ω/Γ, but *a priori* one might imagine that there could be quasiconformal conjugacies supported on the limit set Λ. For geometrically finite Γ this is impossible, since

Theorem 4.24 shows that the space of quasi-isometric deformations of $M := \mathbb{H}^3/\Gamma$ is parameterized by quasiconformal deformations of the boundary. Such deformations with Beltrami differentials supported in Ω are parameterized by the Teichmuller spaces of Ω/Γ. But this leaves open the possibility that there might be Beltrami differentials μ for Γ supported in the limit set.

Sullivan [19] showed this is impossible:

Theorem 5.19 (Sullivan; no invariant line field)**.** *Let Γ be a finitely generated Kleinian group. Then any Γ-invariant Beltrami differential μ vanishes a.e. on the limit set Λ.*

Following McMullen ([14], Thm. 2.10) we give a proof in case the injectivity radius is bounded above by some constant R in the convex core $C(\Gamma) \subset M = \mathbb{H}^3/\Gamma$.

Proof. Given any Γ-invariant Beltrami differential there is a Γ-invariant Beltrami differential with the same argument and support, and constant absolute value 1 on its support, obtained by replacing μ by $\mu/|\mu|$ where it is nonzero. Such a Beltrami differential is called a *line field*, and may be thought of as a measurable section of the real projective tangent bundle on its support.

Let's suppose that there is a Γ-invariant Beltrami differential which does not vanish a.e. on the limit set. Without loss of generality we may suppose μ is a line field supported entirely on the limit set. By the Lebesgue density theorem there is a point of density for μ, i.e. a point $p \in \Lambda$ where $|\mu(z) - \mu(p)| < \epsilon$ on $1 - \epsilon$ of the measure of any sufficiently small round ball centered at p (in some complex chart). By a Möbius transformation we put p at 0 in the upper half-space model in such a way that ∞ is another point in Λ so that the vertical geodesic γ joining them is in the convex hull. We may assume p is not one of the countably many fixed points for Γ, so that the projection of γ to $M := \mathbb{H}^3/\Gamma$ does not eventually stay in the thin part, and enters the thick part infinitely often.

Fix $q \in \gamma$ and let g_i be a sequence of elements of $\mathrm{PSL}(2,\mathbb{C})$ taking γ to itself and taking q to $q_i := g_i q$ converging to p and projecting to the thick part of M. The Kleinian groups $g_i^{-1}\Gamma g_i$ have a subsequence that converges geometrically to some limit Γ_∞, and by Lemma 5.18 the injectivity radius is bounded above by R in $C(\Gamma_\infty)$ and furthermore $\Lambda_\infty = \lim g_i^{-1}\Lambda$ is the entire S^2_∞. Because μ is nearly continuous at p, the pullbacks $g_i^*\mu$ converge (weakly) to a *constant* (i.e. translation invariant) nonzero line field μ_∞ on $\mathbb{C}$. On the other hand, since each $g_i^*\mu$ is $g_i^{-1}\Gamma g_i$-invariant, the limit μ_∞ is Γ_∞ invariant which is absurd, since any isometry fixing μ_∞ must fix infinity. □

5.4. **Pleated surfaces.** Pleating locus, which laminations are realizable, boundary of convex hull, ending laminations.

6. Hyperbolization for Haken manifolds

In this section we give the outline of Thurston's hyperbolization theorem for Haken manifolds. The statement of the full hyperbolization theorem is as follows:

Theorem 6.1 (Hyperbolization). *Let M be a closed oriented 3-manifold or a compact oriented 3-manifold whose boundary are tori. Then the interior of M admits a complete hyperbolic structure with finite volume* unless *one or more of the following conditions holds:*

(1) M contains an essential 2-sphere;
(2) M contains a non-peripheral essential torus or $M = T^2 \times I$; or
(3) $\pi_1(M)$ is finite.

This theorem was conjectured by Thurston in the 1970s, and not proved in full generality until 2001 by Perelman using Ricci flow. We shall give an outline of Perelman's argument in Chapter 7, but well before 2001 Thurston proved the hyperbolization theorem in the special case that M is *Haken*; note that this condition holds automatically if M is irreducible and has nonempty boundary. Recall from Chapter 1 that a Haken manifold is obtained inductively from a collection of balls (really, balls with corners) by repeatedly gluing along essential subsurfaces of the boundary. The inductive argument proceeds by deforming the hyperbolic structure at each stage so that each gluing step can be performed isometrically.

The base step of the induction pertains to the balls with corners (i.e. combinatorial polyhedra) and gives necessary and sufficient conditions for such polyhedra to be realized as hyperbolic polyhedra with all dihedral angles $\pi/2$. This base step is a special case of a theorem of Andreev [3] but was rediscovered (and generalized) by Thurston. It is this special case that we shall refer to as Andreev's Theorem (i.e. Theorem 6.6) and prove in § 6.3.

6.1. **Circle packing.** Before we get to Andreev's Theorem we'll look at the closely related topic of circle packing. A *circle packing* is a collection of round disks in the (round) sphere whose interiors are disjoint. The *nerve* of the packing is the graph (embedded in the sphere) with one vertex for each disk, and one edge for each pair of mutually tangent disks. Conformal automorphisms of the sphere take round disks to round disks, and we think of two circle packings as being isomorphic if they're related by a conformal automorphism of the sphere.

It's natural to wonder, given a graph in the sphere, whether it can be realized as the nerve of some circle packing. Two different circles can't be tangent at more than one point, and we don't consider a circle to be tangent to itself, so the 1-skeleton of the nerve is a graph, embedded in the sphere, without loops or 2-cycles. Any such graph extends to the 1-skeleton of a triangulation τ.

Example 6.2 (triangle). The simplest triangulation of the sphere consists of two triangles glued edge to edge. The 1-skeleton is a 3-cycle, and is realized by three circles of equal radius centered at the vertices of an equilateral triangle; see Figure 16.

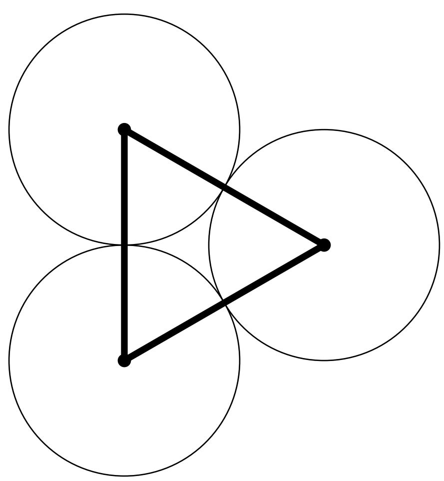

FIGURE 16. Three mutually tangent circles in the sphere can be taken to this configuration by a Möbius transformation.

Any other configuration of three mutually tangent circles is equivalent to this one, by applying a conformal automorphism that takes the three points of tangency to the three midpoints of the equilateral triangle.

Circle packings with these graphs are the most rigid and the most interesting. It turns out that the situation is as nice as possible: for any triangulation of the sphere there is a circle packing, and this circle packing is unique up to isomorphism.

Theorem 6.3 (Circle packing)**.** *Let τ be a triangulation of the sphere. Then there exists a circle packing of the round sphere whose nerve is isomorphic to the one dimensional skeleton of τ. Moreover, any two such circle packings are related by a conformal automorphism of the sphere.*

Lemma 6.4. *It suffices to prove this theorem for triangulations τ for which every 3-cycle bounds a triangle.*

Proof. If τ contains a 3-cycle which is not the boundary of a triangle we can cut along this 3-cycle to produce two (simpler) triangulations τ_1, τ_2 of the sphere. If we can find circle packings realizing τ_1 and τ_2 there is a configuration of three tangent circles in each corresponding to the given 3-cycle. There is a (unique) conformal automorphism taking the triple of circles in τ_1 to the triple of circles in τ_2 (c.f. Example 6.2) and the union gives a circle packing realizing τ. □

6.2. **Ken Stephenson's Miracle Grow Method.** We'll now give Ken Stephenson's miraculous proof of Theorem 6.3; see [18], Chapter 6. Suppose τ is a triangulation of the sphere. Pick a vertex v. The triangles not containing v together give a triangulation of a polygon P whose boundary is the link of v. We can build a circle packing realizing τ by thinking of the round disk associated to v as the *outside* of the unit disk in the plane. Then all the disks associated to vertices in P are on the inside of the unit disk, and the vertices on ∂P are tangent to the unit circle. For example, Figure 17 is a packing corresponding to a triangulation with 400 vertices.

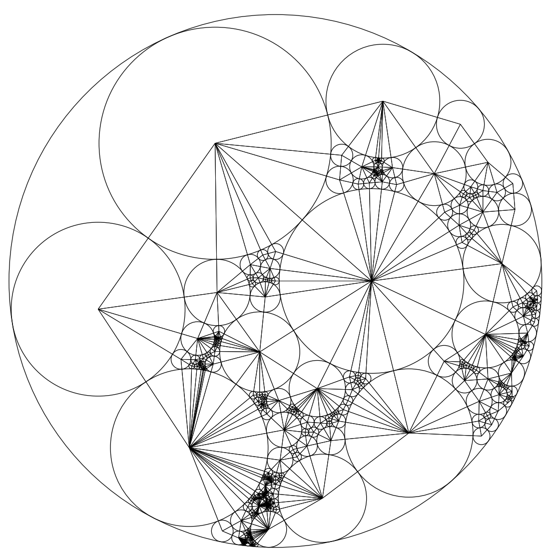

FIGURE 17. A circle packing with 400 circles associated to a triangulated polygon P

Now: identify the unit disk with the hyperbolic plane $\mathbb{H}$. The disks associated to interior vertices of P become round hyperbolic disks, and the disks associated to boundary vertices of P become horoballs. Finding a spherical circle packing for v is the same thing as finding a hyperbolic circle packing for P for which every boundary circle is a horocycle.

If there were an interior edge of P between two (non-adjacent) boundary vertices, this would give a 3-cycle in τ, and we've already seen by Lemma 6.4 that this case splits into two simpler cases. So let's assume by induction that non-adjacent boundary vertices never share an edge.

Let's look for a circle packing for P with horocycles for boundary circles. We can formulate this as a geometric problem. To each vertex v_i of P we associate a hyperbolic length $\ell_i > 0$, where $\ell_i = \infty$ if and only if v_i is in ∂P. Each triangle Δ of the triangulation with vertices v_i, v_j, v_k determines a (possibly semi-ideal) hyperbolic triangle, unique up to isometry, whose vertices are the centers of mutually tangent round (hyperbolic!) disks of radii ℓ_i, ℓ_j, ℓ_k. Note that for v_i a boundary vertex the 'center' of the horoball is the point of tangency on $\partial\mathbb{H}$. At every interior vertex this triangle has an angle determined by these lengths, and we let $\mathcal{L}$ denote the space of length functions $\ell : \{\text{interior vertices}\} \to \mathbb{R}^+$ for which the sum of the angles at every interior vertex is $\ge 2\pi$.

There are a few elementary facts to check about $\mathcal{L}$.

(1) the space $\mathcal{L}$ is compact; in fact, there's an *a priori* upper bound on each finite ℓ_i. That's because for any $n \ge 3$ there's an $r(n)$ so that around a hyperbolic disk of radius $r(n)$ we can fit exactly n tangent horocycles. If n is the valence of an interior vertex v_i, and $\ell_i > r(n)$, then the sum of the interior angles at v_i is necessarily $< 2\pi$.

(2) a function $\ell \in \mathcal{L}$ corresponds to a circle packing if and only if every interior angle sum is 2π. That's because we can lay the triangles out in the hyperbolic plane edge to edge, and there is no holonomy around vertices. By the way, the ideal edges only occur on the boundary, because our polygon P has no interior edges running between adjacent boundary vertices. That's why there's no ambiguity when we say we lay triangles out 'edge to edge'.

Define the *angle excess* to be the sum over all interior vertices of the angle sum minus 2π. This is non-negative, and equal to zero only at an ℓ corresponding to a circle packing. Any ℓ determines a collection of triangles, whose areas sum to $\text{area}(\ell)$, and at a solution $\text{area}(\ell) = (|\partial P| - 2)\pi$ by Gauss–Bonnet, where $|\partial P|$ is the number of vertices in ∂P. Thus the angle excess is equal to $(|\partial P| - 2)\pi - \text{area}(\ell)$. So to make this angle excess go to zero we just have to make the triangles bigger! This is Stephenson's Miracle Grow Method.

More precisely: suppose we have some ℓ for which the angle sum at an interior vertex v_i is strictly bigger than 2π. We claim that by increasing ℓ_i monotonically we can reduce the angle sum at v_i until it is exactly 2π, while the angle sum at the interior vertices adjacent to v_i is increased; and at the same time, the net effect is to decrease the angle excess. This follows from a lemma in hyperbolic trigonometry:

Lemma 6.5 (Angle monotonocity for triangles)**.** *Given three lengths ℓ_i, ℓ_j, ℓ_k form a hyperbolic triangle with side lengths $\ell_j + \ell_k$, $\ell_i + \ell_k$, $\ell_i + \ell_j$ opposite vertices with angles α, β, γ. If we increase ℓ_i while keeping ℓ_j, ℓ_k fixed, then*

$$\frac{\partial \alpha}{\partial \ell_i} < 0, \quad \frac{\partial \beta}{\partial \ell_i} > 0, \quad \frac{\partial \gamma}{\partial \ell_i} > 0 \text{ and } \frac{\partial \text{area}}{\partial \ell_i} > 0$$

We'll prove a generalization of this — Lemma 6.7 — in the next section.

From the lemma the claim follows. We can move around in $\mathcal{L}$, adjusting the lengths monotonically one by one to adjust the angle sum at some vertex to 2π. By compactness and (transfinite) induction any sequence of adjustments eventually takes the angle excess to zero.

The one missing ingredient is that we need to show $\mathcal{L}$ is nonempty. Now, by hypothesis, our polygon P has no edges between nonadjacent boundary vertices. Embed P in the sphere and add edges to the complement to get a triangulation of the sphere τ' with one fewer vertices than τ. Realize this by a circle packing (by induction), put this circle packing in the plane, then rescale it so it fits inside the unit disk. The hyperbolic radii at interior vertices give a function ℓ with angle sum $\geq 2\pi$ at every interior vertex. In fact, the angle sum is 2π at every interior vertex not adjacent to the boundary, and is $> 2\pi$ otherwise. This completes the proof of Theorem 6.3 — except for the uniqueness part.

6.3. **Andreev's Theorem.** Actually, uniqueness in Theorem 6.3 is not hard to prove directly. But there is an elegant proof via Mostow Rigidity that goes to the heart of the hyperbolization theorem.

Suppose we have a circle packing associated to a triangulation τ. Every triangle in τ gives rise to three tangent circles, and there is a unique fourth circle — a so-called 'face circle' that goes through the three points of tangency. Each face circle is perpendicular to the three 'vertex circles' it intersects, so two face circles associated to adjacent faces are tangent. Thus associated to a triangulation τ and a circle packing, we get a bigger collection of circles that are mutually tangent or meet at right angles. See Figure 18.

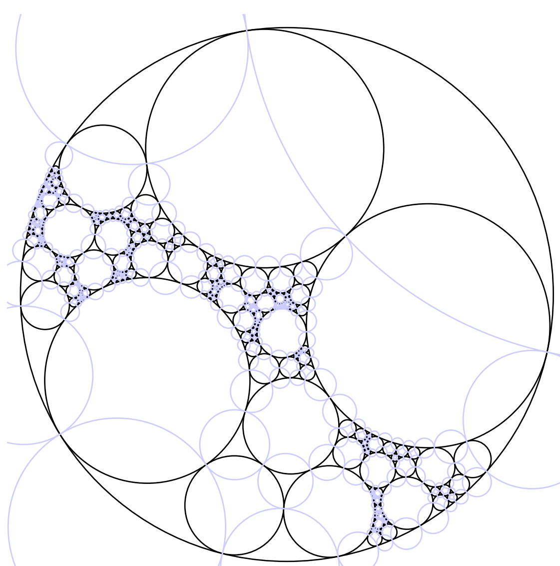

FIGURE 18. Face circles associated to a circle packing

Let's think of this configuration of circles as the boundaries of hyperbolic planes in $\mathbb{H}^3$. When two circles cross at right angles so do the corresponding planes. Thus these planes cut out a finite-sided convex right-angled ideal polyhedron A. The vertices are all 4-valent, and correspond to the points of tangency of the original packing (there is one for every edge of τ). This polyhedron has an ideal triangle for every triangle of τ and an ideal n-gon for every n-valent vertex.

Because A has all right angles and all vertices ideal, the group generated by reflections in the faces of A is discrete and finite covolume. A different circle packing associated to τ would give a different, but abstractly isomorphic, Kleinian group. Mostow rigidity says these groups are conjugate, and so are the circle packings. This proves uniqueness.

Andreev's Theorem gives necessary and sufficient conditions to realize a right angled semi-ideal finite-volume polyhedron A with prescribed combinatorics. The 3-valent vertices of A are finite, and the 4-valent ones are ideal. There should be no vertices of valence > 4 because a right-angled n-gon for $n > 4$ is hyperbolic, and A couldn't have finite volume. First let's discuss some necessary conditions.

As above, the group generated by reflections in the faces of a hyperbolic realization of A will be discrete and have finite covolume, and therefore will contain a finite index torsion-free subgroup, whose quotient will be a hyperbolic manifold M_A. One rather explicit and canonical way to produce M_A as a topological manifold directly from the combinatorial polyhedron A is as follows. Let $\mathcal{F}$ denote the (finite) set of faces of A, and let $2^{\mathcal{F}}$ denote the set of functions from $\mathcal{F}$ to the set $\{\pm 1\}$. We take one copy A_f of A for each $f \in 2^{\mathcal{F}}$, and glue A_f and $A_{f'}$ along the face $F \in \mathcal{F}$ if and only if the functions f and f' agree everywhere *except* F. Along every edge e four copies of A are glued, corresponding to sets of four functions that differ only along the two faces adjacent to e in A; likewise at every finite vertex v eight copies of A are glued, corresponding to the sets of eight functions that differ only along the three faces adjacent to v in A. Notice that the 4-valent vertices of A become cusps of M_A whose links are tori; in particular we should remove these vertices of A before doubling to ensure that M_A is a topological manifold. If the polyhedron A is realizable, then M_A admits a complete finite volume hyperbolic structure.

Any n-cycle γ in the dual graph to the 1-skeleton of A bounds a (combinatorial) n-gon P_γ in A covered by a surface S_γ in M_A tiled by copies of P_γ, four meeting at every vertex. If γ is a 3-cycle decomposing A into two pieces $A_\gamma^\pm$, then P_γ is a triangle. Any closed surface tiled by triangles meeting four at a vertex is (virtually) a 2-sphere. By the first bullet of Theorem 6.1, in order for M_A to be hyperbolic, this 2-sphere should bound a ball in M_A, and this ball will be stabilized by the subgroup of Γ generated by reflections in the sides of one of $A_\gamma^\pm$ (without loss of generality, say A_γ^+); in particular this group is *finite*. This is impossible unless the faces of A_γ^+ share a single common (finite) vertex — equivalently, unless γ is the link of a 3-valent vertex. A 3-cycle which is not the link of a 3-valent vertex is *essential*.

Similarly, if γ is a 4-cycle, S_γ is tiled by squares, four meeting at every vertex, and is therefore a torus or Klein bottle. Again, this implies that the subgroup of Γ generated by reflections in the sides of A_γ^+ is virtually a quotient of $\mathbb{Z}^2$, so that γ must be either the link of an edge between adjacent 3-valent vertices or the link of a single 4-valent vertex. A 4-cycle which is not the link of a 4-valent vertex is *essential*. If every 4-cycle is inessential, then every torus in M_A will be boundary parallel; this is compatible with the second bullet of Theorem 6.1 unless M_A is the interior of $T^2 \times I$ which can only occur if A is a *square pyramid*; i.e. a polyhedron with one square and four triangular faces.

Finally there is one sporadic nonexample: in a (combinatorial) tetrahedron every 3-cycle in the dual graph is the link of a 3-valent vertex, and every 4-cycle in the dual graph is the link of an edge joining 3-valent vertices (actually, two such edges). But there is no right-angled hyperbolic tetrahedron: the manifold M_A constructed topologically from A as above is S^3, made from 16 right-angled *spherical* tetrahedra which are the intersection of the round unit S^3 in $\mathbb{R}^4$ with the coordinate orthants. This is the only possibility with $\pi_1(M_A)$ finite, parallel to the third bullet of Theorem 6.1.

These necessary conditions are *sufficient*:

Theorem 6.6 (Andreev's Theorem)**.** *Let A be a combinatorial 3-dimensional polyhedron whose underlying space is homeomorphic to a ball, and which has vertices of valence* 3 *or* 4*. Then A may be realized by a (combinatorially equivalent) semi-ideal right-angled hyperbolic polyhedron whose 3-valent vertices are finite and whose 4-valent vertices are ideal* unless *one or more of the following conditions holds:*

(1) some 3-cycle in the dual complex does not link a 3-valent vertex;
(2) some 4-cycle in the dual complex does not link a 4-valent vertex or an edge joining 3-valent vertices and A is not a square pyramid; or
(3) A is a (combinatorial) tetrahedron.

Proof. We shall prove the theorem in the simpler case that A has all vertices of valence 3, although from time to time we briefly indicate how the argument should be modified to handle 4-valent vertices.

Suppose we could realize A as a right-angled hyperbolic polyhedron. The faces would be tangent to totally geodesic hyperbolic planes, each asymptotic to a round circle in the sphere at infinity. Two faces meeting along an edge with dihedral angle α corresponds to a pair of circles intersecting at angle α. So to realize A is the same thing as finding a configuration of round circles in the sphere, one for each vertex of the dual polyhedron B, where two circles intersect (and at right angles) if the corresponding vertices share an edge. Let's call this (by abuse of notation) a *right-angled circle packing*. More generally we could talk about a ν circle packing where pairs of circles corresponding to edges intersect at a prescribed common angle $\nu < \pi/2$. The idea will be to find a $\pi/2$-packing by analytic continuation of a family of ν-packings, starting at $\nu = 0$ (i.e. an ordinary circle packing) and increasing ν monotonically to $\pi/2$. When A has all vertices of valence 3 this is the whole story; when A has vertices of valence 4 we need to insist that each chain of four circles in a ν-packing dual to a 4-valent vertex are simultaneously orthogonal to a single circle; this extra 'rigidity' condition ensures that the ν-packing is unique.

How can one find a ν-packing? We can try to adapt Ken Stephenson's Miracle Grow Method. And — it works! Sort of. Let τ denote the triangulation of the sphere associated to ∂B. Pick one vertex v of τ and let P be the complementary triangulated polygon. We associate v to the circle at infinity S^1_∞ of the hyperbolic plane. Each vertex w of ∂P will give us a circle γ_w making angles of ν with S^1_∞; for $\nu > 0$ this is a curve of constant distance to a hyperbolic geodesic. Each interior vertex v_i will give us a hyperbolic circle γ_i of radius ℓ_i

For each triangle of τ with vertices v_i, v_j, v_k we get three circles of radii ℓ_i, ℓ_j, ℓ_k. For any fixed angle $0 \le \nu < \pi$ there is a unique way to place three such circles $\gamma_i, \gamma_j, \gamma_k$ (up to isometry) so that they all meet each other with angle ν. Drawing lines from the three centers to the three outermost points of intersection we get a convex hyperbolic hexagon with edge lengths (in cyclic order) $\ell_i, \ell_i, \ell_j, \ell_j, \ell_k, \ell_k$ and angles are $\alpha, \nu, \beta, \nu, \gamma, \nu$ where α is the angle between the two sides of length ℓ_i, and so on. For fixed ν, each of the angles α, β, γ is a function of all three ℓ_i, ℓ_j, ℓ_k. We denote the isometry type of this hexagon by $H(\nu; \ell_i, \ell_j, \ell_k)$. An assignment corresponds to a ν circle packing if and only if the sum of these angles at every interior v_i is exactly 2π. Let $\mathcal{L}_\nu$ be the space of length assignments for which the sum of angles at every interior vertex v_i is at least 2π.

The promised generalization of Lemma 6.5 (which corresponds to the case $\nu = 0$) is as follows:

Lemma 6.7 (Angle monotonicity for hexagons)**.** *For fixed ν, ℓ_j, ℓ_k consider a family of hyperbolic hexagons $H(\nu; \ell_i, \ell_j, \ell_k)$ varying with ℓ_i. For each hexagon in the family let α, β, γ be the angles whose adjacent edges have lengths ℓ_i, ℓ_j, ℓ_k. Then*

$$\frac{\partial \alpha}{\partial \ell_i} < 0, \quad \frac{\partial \beta}{\partial \ell_i} > 0, \quad \frac{\partial \gamma}{\partial \ell_i} > 0 \text{ and } \frac{\partial\, area(H)}{\partial \ell_i} > 0$$

Proof. In the Poincaré disk model the circles γ_j, γ_k associated to v_j, v_k can be placed so that their *Euclidean* radii are equal. Thus the *Euclidean* center of γ_i will be on the perpendicular bisector of the edge from the Euclidean centers of γ_j and γ_k, and the point at which γ_i meets γ_j and γ_k moves monotonically as a function of its Euclidean radius; this proves $\partial\beta/\partial\ell_i > 0$ and $\partial\gamma/\partial\ell_i > 0$. Since these points of intersection move monotonically away from each other, each hexagon in the family is strictly contained in those with bigger ℓ_i. Thus $\partial \text{area}(H)/\partial\ell_i > 0$. Now Gauss–Bonnet gives $\partial\alpha/\partial\ell_i < 0$. □

If A has 4-valent vertices, we must modify the Miracle Grow method slightly. 4-valent vertices of A correspond to *quadrilateral faces* in the 'triangulation' τ. We may modify τ to τ' by subdividing each such quadrilateral into four triangles, all meeting at a new central vertex. We want a circle packing associated to this configuration which is a 'mixed' $\nu, \pi/2$-packing in the sense that circles meet at angles ν along edges of τ, and at angles $\pi/2$ along edges of $\tau' - \tau$. The necessary modifications to the method are not difficult to figure out and we omit the details.

If there is a ν-packing, let A_ν denote the polyhedron obtained as the intersection of half-spaces in $\mathbb{H}^3$ bounded by the planes bounding the circles of the packing. By construction the dihedral angles are all equal to ν. If $\nu \le 60°$ then A_ν will not be compact, and if $\nu < 60°$ then A_ν will have infinite volume, since three hyperbolic planes that intersect mutally at angles of ν will not have a common point of intersection unless there is a spherical triangle (the link of the common point) with all angles equal to ν. This may be rectified as follows.

Let's suppose first that A has only 3-valent vertices. Three hyperbolic planes that intersect mutually at an angle $\nu < 60°$ are all mutually orthogonal to a unique plane that we call an *orthoplane*. If we intersect A_ν with the hyperbolic half-spaces bounded by these orthoplanes, the result will be a smaller hyperbolic polyhedron $\hat{A}_\nu \subset A_\nu$. When $0 < \nu < 60°$, $\hat{A}_\nu$ is (combinatorially) a *truncated* polyhedron obtained from A by cutting off a little tetrahedral neighborhood of each vertex to produce a new triangular face. When $\nu = 0$, the triangular faces centered at adjacent vertices of A collide to create an (ideal) 4-valent vertex of $\hat{A}_\nu$. When $\nu > 0$, $\hat{A}_\nu$ has an edge with dihedral angle ν for every edge of A, together with a hyperbolic triangle with dihedral angles $\pi/2$ along the edges and internal angles ν for every vertex of A. As ν increases to $60°$ the hyperbolic triangles shrink to ideal vertices. As ν increases past $60°$ these ideal vertices become finite vertices and $\hat{A}_\nu = A_\nu$. In every case $\hat{A}_\nu$ has finite volume. Actually we have already encountered one special case of this construction: taking $\nu = 0$ corresponds to an honest circle packing, and in this case the orthoplanes are the ones that bound the *face circles* associated to the circle packing; see Figure 18.

In case A has 4-valent vertices we modify the construction as follow. Remember that we insisted (by fiat) that when there are 4-valent vertices, our ν-packings should have the additional property that chains of 4 circles linking a 4-valent vertex of A should be mutually orthogonal to a single circle. This circle bounds an orthoplane exactly as in the 3-valent case, and these orthoplanes truncate 4-valent vertices of A for every $\nu < 90°$. Figure 19 shows an example of A and $\hat{A}_\nu$ for $\nu = 0$, $30°$, $60°$ and $90°$.

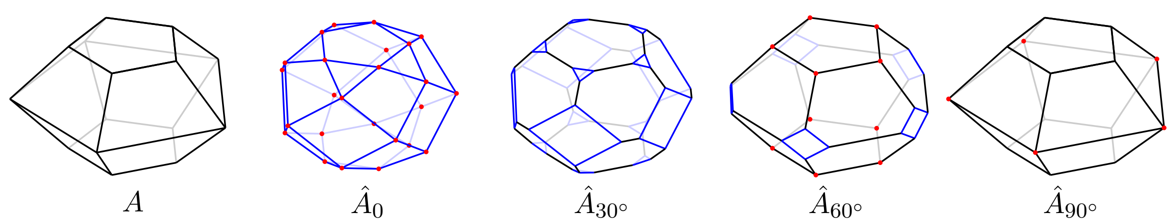

FIGURE 19. The combinatorial polyhedron A and the geometric polyhedra $\hat{A}_\nu$ for $\nu = 0$, $30°$, $60°$, $90°$. Ideal vertices are in red, right-angled edges are in blue, and 'ordinary' edges with dihedral angle ν for $0 < \nu < 90°$ are in black.

We shall complete the proof of Andreev's Theorem by a continuity argument. Let $U \subset [0, \pi/2]$ be the set of ν for which the polyhedron $\hat{A}_\nu$ may be realized. We have already shown that U contains 0. We will show that U is open and closed, and therefore $U = [0, \pi/2]$. Then $\hat{A}_{\pi/2} = A_{\pi/2}$ is the desired realization of A.

The first step is to show that if there is a ν-packing (and therefore an associated point in $\mathcal{L}_\nu$ with zero angle excess) then we may find a nearby point in $\mathcal{L}_\nu$ where all interior angle excesses are *strictly positive*. To see this, we realize the ν-packing in the Poincaré disk model, think of the result as a ν-packing inside a Euclidean unit disk with boundary S^1_∞, shrink the Euclidean circles (but not S^1_∞) homothetically towards the center, then reinterpret the result as a collection of circles in the Poincaré disk. If we now grow the boundary circles until they make angle ν with S^1_∞ the resulting collection has strictly positive interior angle excess at every vertex adjacent to the boundary. We may then iteratively shrink interior circles by smaller and smaller amounts until *every* interior vertex has a positive angle excess. Since the property of a length assignment of having positive angle excess everywhere is open (as a function of ν), we see that $\mathcal{L}_{\nu'}$ is nonempty for all ν' sufficiently close to ν.

Once we know $\mathcal{L}_{\nu'}$ is nonempty, we will obtain a ν'-packing providing $\mathcal{L}_{\nu'}$ is compact. For $\nu' < 90°$ compactness holds for the same argument as when $\nu' = 0$; however there is a subtlety to go from a length assignment $\ell_{\nu'}$ to an honest geometric polyhedron: at certain specific values of ν', under topological conditions on the packing the radius of some circle in the packing might go to zero. This happens at $\nu' = 60°$ if there are essential 3-cycles. Conversely, if there are essential 4-cycles at $\nu' = 90°$ the radius of some circle might go to infinity; see Figure 20. However at any other value $\mathcal{L}_{\nu'}$ will be compact and a length assignment will have strictly positive internal radii, and correspond to an 'honest' geometric polyhedron $\hat{A}_{\nu'}$. Thus (whether there are essential 3- or 4-cycles or not) U is open.

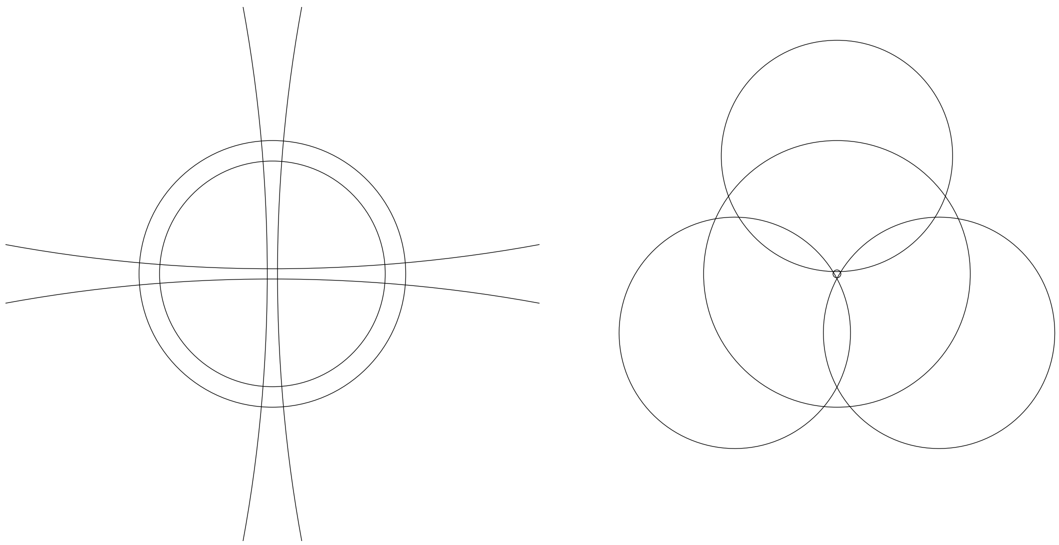

FIGURE 20. An ordinary circle with $n > 4$ neighbors in a ν'-packing has *a priori* positive upper and lower bounds on its radius when $\nu' \leq 90°$. If $n = 4$ there is a positive upper bound for $\nu' < 90°$ and for $n = 3$ there is a positive lower bound for $\nu' < 60°$.

To show that U is closed we consider when $\hat{A}_\nu$ can fail to have a limit. Fix ν', and suppose $\hat{A}_\nu$ exists for all $\nu < \nu'$. Each polyhedron $\hat{A}_\nu$ is cut out by n totally geodesic hyperbolic planes for n equal to the number of faces of A, plus the number of vertices cut off by orthoplanes (which is bounded by the number of vertices of A but might depend on ν). We may always extract a Hausdorff limit X of the $\hat{A}_\nu$ for some sequence of $\nu \to \nu'$. The limit X will be convex and cut out by $\leq n$ planes. The volume of $\hat{A}_\nu$ is monotone decreasing as a function of ν by Schläffli's formula (i.e. Theorem 3.9) and therefore volume$(X) <$ volume$(\hat{A}_\nu)$. There are two ways that X can fail to realize $\hat{A}_{\nu'}$: either the distance between some pair of planes can go to infinity, or the length of some edge can go to zero (or both simultaneously!)

The first case can, and definitely does happen: as $\nu \to 60°$ the orthoplanes will recede to infinity, and in the limit their associated triangles in $\hat{A}_{\nu'}$ will degenerate to ideal vertices. In general, if the distance between two supporting planes goes to infinity, we may join these planes by a mutually perpendicular geodesic segment γ. Since X is convex, any plane H orthogonal to γ will intersect $\hat{A}_\nu$ in a convex polygon, and by the coarea formula and the upper bound on volume$(\hat{A}_\nu)$ the areas of these polygons must get arbitrarily small. If we choose H sufficiently far from a vertex of $\hat{A}_\nu$ (which is possible if γ is very long, since there is an *a priori* bound on the number of vertices of $\hat{A}_\nu$), the intersecting polygon will have interior angles arbitrarily close to $\nu \leq 90°$. Thus by Gauss–Bonnet there are only two possibilities: either the polygon is a triangle, and $\nu \to 60°$ or the polygon is a quadrilateral and $\nu \to 90°$. Except for the degenerating triangles or quadrilaterals contained in orthoplanes, such a triangle or quadrilateral will be dual to an essential 3- or 4-cycle.

If the length of some edge goes to zero, then if $\nu' < 90°$ the only possibility is that the entire polyhedron shrinks to zero size. If we rescale the metric so that the diameters of the shrinking polyhedra are equal to 1 we will obtain in the limit a nontrivial *Euclidean* polyhedron, combinatorially equivalent to A, with dihedral angles ν'. Since $\nu' < 90°$ we may approximate the Euclidean polyhedron by a *spherical* polyhedron with slightly larger dihedral angles (but still smaller than 90°) and perform the doubling construction to obtain a singular spherical metric on M_A. This metric may be smoothed along the singularities to obtain a Riemannian metric on M_A with strictly positive sectional curvatures. A closed manifold with strictly positive sectional curvatures has finite fundamental group.

The condition that M_A has finite fundamental group imposes very stringent conditions on A. If A has a 4-valent vertex, then M_A has at least one torus boundary component, so $H_1(M_A)$ is infinite for the usual reason. If A contains two faces F_1, F_2 that are not adjacent but both contain the endpoints of some edge e of a third face, then in M_A there is a connected surface S projecting to F_1 and a circle γ projecting to e that intersect transversely in one point; in particular $H_1(M_A)$ (and therefore also $\pi_1(M_A)$) is infinite. Thus if $\pi_1(M_A)$ is finite, all vertices are 3-valent and every two faces are adjacent; in particular, A is a tetrahedron. Indeed, in this case $\hat{A}_\nu$ shrinks to a point as $\nu \to \nu' \sim 70.529°$ (i.e. $\arccos(1/3)$).

If the length of some edge goes to zero at $\nu' = 90°$ then $\hat{A}_\nu$ might *collapse* to a polyhedron of lower dimension. If the limiting polyhedron is 2-dimensional it is a right-angled polygon, and the link of an edge of this polygon is an essential 4-cycle in A. Notice in this case if A has no 4-valent vertices then it is a prism, i.e. the product of a polygon with an interval. If A has 4-valent vertices then $\hat{A}_\nu$ will be a prism for ν close to 90°. If the limit is 1-dimensional it is a segment — finite, in which case A is a combinatorial cube which contains an essential 4-cycle — or infinite, in which case A is a square pyramid. If the limit is 0-dimensional, A is again a combinatorial cube. This completes the proof of Andreev's Theorem. □

6.4. **The Skinning map.** Now let's return to the hyperbolization theorem. Suppose M is a Haken manifold, closed for simplicity, and $S \subset M$ is a two-sided essential surface. Let M' be the result of cutting M open along S, so that $\partial M'$ consists of two copies of S with either orientation. Let's suppose further that we have already found a geometrically finite hyperbolic structure on M'. Once we have found one geometrically finite hyperbolic structure on M', all the others are parameterized by the Teichmüller space of the boundary, i.e. by $\mathcal{T}(S) \times \mathcal{T}(\bar{S})$. If we can deform the hyperbolic structure on M' so that gluing up the two boundary components can be performed isometrically, then we will obtain a hyperbolic structure on M. Thurston constructs a self-map σ of $\mathcal{T}(S) \times \mathcal{T}(\bar{S})$ (the *skinning map*) with the property that we can solve the deformation problem if and only if σ has a fixed point.

6.5. **Bounded image theorem.**

6.6. **Only windows break.**

6.7. **Orbifold trick and reduction to the last step.**

6.8. **Double limit theorem.**

6.9. **Fibered case.**

7. Tameness

7.1. **Geometrically infinite manifolds.**

7.2. **Bonahon's exiting sequences.**

7.3. **Shrinkwrapping.**

7.4. **Ahlfors' Conjecture.**

8. Ending laminations

9. Acknowledgments

Danny Calegari was supported by NSF grant DMS 1405466.

References

[1] I. Agol, *The minimal volume orientable hyperbolic 2-cusped 3-manifolds*, Proc. AMS **138** (2010), no. 10, 3723–3732

[2] L. Ahlfors, *Lectures on Quasiconformal Mappings*, D. Van Nostrand, Princeton, New Jersey, 1966

[3] E. M. Andreev, *On convex polyhedra in Lobacevskii spaces (English Translation)*, Math. USSR Sbornik, **10** (1970), 413–440

[4] C. Cao and R. Meyerhoff, *The orientable cusped hyperbolic 3-manifolds of minimum volume*, Invent. Math. **146** (2001), no. 3, 451–478

[5] A. Douady and C. Earle, *Conformally natural extension of homeomorphisms of the circle*, Acta Math. **157** 23–48

[6] D. Epstein and C. Petronio, *An exposition of Poincaré's polyhedron theorem*, L'Ens. Math. **40** (1994), 113–170

[7] Fathi, Laudenbach and Poenaru, *Travaux de Thurston*

[8] D. Gabai, R. Meyerhoff and P. Milley, *Minimum volume cusped hyperbolic 3-manifolds*, Jour. AMS **22** (2009), no. 4, 1157–1215

[9] U. Haagerup and H. Munkholm, *Simplices of maximal volume in hyperbolic n-space* Acta Math. **147** (1981), no. 1-2, 1–11

[10] J. Harer and R. Penner, *Combinatorics of train tracks*, Princeton

[11] J. Hubbard, *Teichmüller theory*, volume 1

[12] T. Kuessner, *Proportionality principle for cusped manifolds*, Arch. Math. **43** (2007), 485–490

[13] A. Marden and B. Rodin, *On Thurston's formulation and proof of Andreev's theorem*, Comp. Meth. and Function Thy. Proceedings, Valparaiso, 1989, Springer Lecture Notes in Mathematics **1435**, 103–115, Springer, Berlin, 1990

[14] C. McMullen, *Renormalization and 3-manifolds which fiber over the circle*, Ann. Math. Studies 142, Princeton University Press, Princeton, New Jersey, 1996

[15] C. Morrey, *On the solutions of quasi-linear elliptic partial differential equations*, Trans. AMS **43** (1938), 126–166

[16] W. Neumann and D. Zagier, *Volumes of hyperbolic three-manifolds*, Topology **24** (1985), no. 3, 307–332
[17] I. Rivin, *A characterization of ideal polyhedra in hyperbolic 3-space*, Ann. of Math. (2) **143** (1996), no. 1, 51–70
[18] K. Stephenson, *Introduction to Circle Packing: The Theory of Discrete Analytic Functions*, Cambridge University Press, Cambridge, 2005.
[19] D. Sullivan, *On the ergodic theory at infinity of an arbitrary discrete group of hyperbolic motions*, Riemann surfaces and related topics, Ann. Math. Studies 97, Princeton University Press, Princeton, New Jersey, 1981
[20] W. Thurston, *The Geometry and Topology of 3-Manifolds*, a.k.a "Thurston's Notes"; available from the MSRI at `http://www.msri.org/publications/books/gt3m`
[21] W. Thurston, *On the geometry and dynamics of diffeomorphisms of surfaces*, Bull. AMS **19** (1988), no. 2, 417–431
[22] W. Thurston, *Hyperbolic structures on 3-manifolds, I: Deformation of acylindrical manifolds*, Ann. Math. (1986)
[23] W. Thurston, *Hyperbolic structures on 3-manifolds, II: Surface groups and 3-manifolds which fiber over the circle*, `https://arxiv.org/abs/math/9801045`
[24] S. Wolpert *On the symplectic geometry of deformations of a hyperbolic surface*, Ann. Math. **117** (1983), 207–234

University of Chicago, Chicago, Ill 60637 USA
Email address: `dannyc@math.uchicago.edu`