NOTES ON KLEINIAN GROUPS

DANNY CALEGARI

Abstract. These are notes on Kleinian groups. These notes follow a course given at the University of Chicago in Spring 2015.

Contents

1. Models of hyperbolic space 1
2. Building hyperbolic manifolds 11
3. Rigidity and the thick-thin decomposition 20
4. Quasiconformal deformations and Teichmüller theory 32
5. Hyperbolization for Haken manifolds 46
6. Tameness 46
7. Ending laminations 46
8. Acknowledgments 46
References 47

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

\[ f''(\theta) + \lambda f(\theta) = 0 \]  

for some real constant \( \lambda \), where \( f \) is a smooth real-valued function of a real variable \( \theta \). The equation is 2nd order and linear so the space of solutions \( V_\lambda \) is a real vector space of
dimension 2, and we may choose a basis of solutions $c(\theta), s(\theta)$ normalized so that if $W(\theta)$
denotes the Wronskian matrix
\begin{equation}
W(\theta) := \begin{pmatrix} c(\theta) & c'(\theta) \\ s(\theta) & s'(\theta) \end{pmatrix}
\end{equation}
then $W(0)$ is the identity matrix.

Since the equation is autonomous, translations of the $\theta$ coordinate induce symmetries of $V_\lambda$. That is, there is an action of (the additive group) $\mathbb{R}$ on $V_\lambda$ 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{equation}
\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}
\end{equation}

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 $\textit{hyperbolic length}$ on the hyperboloid.

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{projection.png}
\caption{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.}
\end{figure}

1.1.2. $\textit{Projection to the tangent}$. Linear projection from the origin to the tangent line at $(1, 0)$ takes the coordinate $\theta$ to the $\textit{projective}$ 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 \( \theta \)-line becomes the addition law for ordinary and hyperbolic tangent:

\[
\tan(\alpha + \beta) = \frac{\tan(\alpha) + \tan(\beta)}{1 - \tan(\alpha) \tan(\beta)}; \quad \tanh(\alpha + \beta) = \frac{\tanh(\alpha) + \tanh(\beta)}{1 + \tanh(\alpha) \tanh(\beta)}
\]

1.1.3. **Stereographic projection.** Stereographic linear projection from \((-1, 0)\) to the \(y\)-axis takes the coordinate \(\theta\) 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 \(\rho\). The addition law for translations on the \(\theta\)-line expressed in terms of \(\rho_E\) for the circle and \(\rho_M\) for the hyperboloid, are

\[
\rho_E(\alpha + \beta) = \frac{\rho_E(\alpha) + \rho_E(\beta)}{1 - \rho_E(\alpha) \rho_E(\beta)}; \quad \rho_M(\alpha + \beta) = \frac{\rho_M(\alpha) + \rho_M(\beta)}{1 + \rho_M(\alpha) \rho_M(\beta)}
\]

The only solutions to these functional equations are of the form \(\tan(\lambda \theta)\) and \(\tanh(\lambda \theta)\) for constants \(\lambda\), 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

\[
Q_E = z^2 + \sum x_i^2 \quad \text{and} \quad Q_M = -z^2 + \sum x_i^2
\]

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 \(O(Q)\) denote the group of linear transformations of \(\mathbb{R}^{n+1}\) preserving the form \(Q\).

In terms of formulae, a matrix \(M\) is in \(O(Q)\) if \((Mv)^T Q(Mv) = v^T Q v\) for all vectors \(v\); or equivalently, \(M^T Q M = Q\). Denote by \(SO^+(Q)\) the connected component of the identity in \(O(Q)\). 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 \(SO(n+1)\) and \(SO^+(n, 1)\) for \(SO^+(Q)\) if we want to stress the signature and the dependence on the dimension \(n\).

**Example 1.1.** If \(n = 1\) then \(SO^+(Q)\) 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 \(SO^+(Q)\) preserves \(X\), and acts transitively with point stabilizers isomorphic to \(SO(n, \mathbb{R})\).

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

Denote by \(p\) the point \(p = (0, \cdots, 0, 1)\). Then \(p \in X\) and its stabilizer acts faithfully on \(T_p X\) 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 \(SO(n, \mathbb{R})\), 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 \( \text{SO}(n, \mathbb{R}) \) (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 \( \tau \). Then the matrix

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

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

Denote by \( A_H \) the subgroup of \( \text{SO}^+(Q_M) \) consisting of matrices \( A(\tau) \) as above, and by \( A_S \) the subgroup of \( \text{SO}(Q_E) \) consisting of matrices \( I_{n-1} \oplus W(\theta) \), and denote either subgroup by \( A \). Similarly, in either case denote by \( K \) the subgroup \( \text{SO}(n, \mathbb{R}) \) 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 (KA\(K\) decomposition).** Every matrix in \( \text{SO}^+(Q) \) 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 \( \text{SO}(n-1, \mathbb{R}) \) unless \( a \) is trivial).

**Proof.** Let \( g \in \text{SO}^+(Q) \) 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 \( \text{SO}(n-1, \mathbb{R}) \). \( \square \)

It is useful to spell out the relationship between matrix entries in \( \text{SO}^+(Q) \) 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 \( \text{SO}^+(Q) \) if and only if the last column is a vector \( v \) on \( X \), and the first \( n \) columns are an (oriented) orthonormal basis for \( T_p X \).

**Proof.** This is true for the identity matrix, and it is therefore true for all \( M \) because \( \text{SO}^+(Q) \) 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 \( T_p X \) is positive definite, it inherits the structure of a Riemannian manifold. The group \( \text{SO}^+(Q) \) 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

\[
\cos(\angle(a, b)) = \frac{a^T Q b}{\|a\|\|b\|}
\]

Since the action of \( \text{SO}^+(Q) \) 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 \( g \in \text{SO}^+(Q) \) 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 \( \text{SO}(n - 1, \mathbb{R}) \) fixes precisely this curve pointwise, so it must be totally geodesic. In particular, in this example, \( d(v, w) = \tau \), so that

\[
c(d(v, w)) = \frac{v^TQw}{\|v\|\|w\|}
\]

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 \( \text{SO}^+(Q) \), 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

\[
\cosh(d(v, w)) = \frac{v^TQw}{\|v\|\|w\|}
\]

1.2.3. Sine and cosine rule. Three points \( A, B, C \) on \( X \) span a geodesic triangle with angles \( \alpha, \beta, \gamma \) and lengths \( a, b, c \) (where \( a \) is the length of the edge opposite the angle \( \alpha \) 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 \( u \cdot v := u^TQv \) 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 \( \sin, \cosh \) 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

\[
\cos(a) = \cosh(b)\cos(c) + \sin(b)\sin(c)\cos(\alpha)
\]

and in hyperbolic geometry this gives

\[
\cosh(a) = \cosh(b)\cosh(c) - \sinh(b)\sinh(c)\cos(\alpha)
\]

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 \( \text{SO}^+(Q) \) 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

\[
\frac{\sin(\alpha)}{\sin(a)} = \frac{\sin(\beta)}{\sin(b)} = \frac{\sin(\gamma)}{\sin(c)}
\]
and in hyperbolic geometry this gives
\[
\frac{\sin(\alpha)}{\sinh(a)} = \frac{\sin(\beta)}{\sinh(b)} = \frac{\sin(\gamma)}{\sinh(c)}
\]

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 \( \SO^+(Q) \) acts transitively on the unit tangent bundle of \( X \), every geodesic in \( X \) is the intersection of \( X \) with a 2-plane \( \pi \); 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 \( \SO(n-1) \times A \) which is equal to \( \SO(n-1) \times \SO^+(1,1) \) or \( \SO(n-1) \times \SO(2) \).

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 \( \SO(n-k) \times \SO^+(k,1) \) or \( \SO(n-1) \times \SO(k+1) \).

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 \( \SO^+(Q_H) \) 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 \( \PSL(2, \mathbb{R}) \) in \( \PSL(3, \mathbb{R}) \) obtained by projectivizing \( S^2V \), the symmetric square of the standard representation \( V \) of \( \SL(2, \mathbb{R}) \). This exceptional case is discussed again in § 1.3.2.

If \( \ell \) is a projective line over any field, a projective automorphism of \( \ell \) 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 \( \lambda \) 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 \( \ell \) in \( B \) containing \( p \) and \( q \), and intersecting \( \partial B \) 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:
\[
d_K(p, q) = \frac{1}{2} \log \frac{(q - \ell(0))(\ell(1) - p)}{(p - \ell(0))(\ell(1) - q)}
\]
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 \( \theta \), we obtain

\[
\theta := d_K(p, q) = \frac{1}{2} \log \frac{1 + r}{1 - r}
\]

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. \textit{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 \textit{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 \( \partial B \).

![Figure 2. An ideal triangle in various models.](image)

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 \( \partial B \).

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 \( \text{SO}(n, \mathbb{R}) \) (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. 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
The easiest way to compute hyperbolic distances between points in $B$ in the Poincaré model is to project back to the hyperboloid by

\[(1.15) \quad \left( x_1, \ldots, x_n, 0 \right) \rightarrow \left( \frac{2x_1}{1 - \sum x_i^2}, \ldots, \frac{2x_n}{1 - \sum x_i^2}, \frac{1 + \sum x_i^2}{1 - \sum x_i^2} \right)\]

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

\[(1.16) \quad \theta := d_P(p, q) = \log \frac{1 + r}{1 - r}\]

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 $q' := (0, \ldots, 0, \sinh(\theta), \cosh(\theta))$ on the hyperboloid. By Lemma 1.4, one orthonormal frame at $q'$ is given by vectors

\[v_1 := (1, 0, \ldots, 0), \quad v_2 := (0, 1, 0, \ldots, 0), \quad \ldots, \quad v_n := (0, \ldots, 0, \cosh(\theta), \sinh(\theta))\]

and we deduce that a hyperbolic circle with radius $\theta$ 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 $\pi$ 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

\[(1.17) \quad ds_P = \frac{2ds_E}{1 - r^2}\]

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 $x_1, \ldots, x_{n-1}, z$ 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

\[(1.18) \quad ds_P = \frac{ds_E}{z}\]

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 $\infty$; 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 \( \text{SO}^+(Q) \) 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 \( K = \text{SO}(n; \mathbb{R}) \) 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 \( \text{SO}^+(Q) \) 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. \( \square \)

### 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 \( \text{PGL}(2, \mathbb{C}) = \text{PSL}(2, \mathbb{C}) \), acting projectively by

\[
\begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot z = \frac{az + b}{cz + d}
\]

The subgroup fixing the unit circle is \( \text{PSU}(1, 1) \), whose elements are represented (uniquely up to sign) by matrices of the form \( \begin{pmatrix} \alpha & \beta \\ \overline{\beta} & \overline{\alpha} \end{pmatrix} \) with \(|\alpha|^2 - |\beta|^2 = 1\). The subgroup fixing the real line is \( \text{PSL}(2, \mathbb{R}) \), whose elements are represented (uniquely up to sign) by real matrices of the form \( \begin{pmatrix} a & b \\ c & d \end{pmatrix} \) with \( ad - bc = 1 \). These subgroups are conjugate in \( \text{PSL}(2, \mathbb{C}) \), 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 \( \text{SL}(2, \mathbb{C}) \) which is unique up to multiplication by \(-1\).

1. If \( |\text{tr}(g)| < 2 \) then \( g \) is elliptic. It fixes a unique point in the interior of the hyperbolic plane, and acts as a rotation through angle \( \alpha \) where \( \cos(\alpha/2) = \text{tr}(g)/2 \).
2. If \( |\text{tr}(g)| = 2 \) 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 \( |\text{tr}(g)| > 2 \) 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 \( \cosh(l/2) = \text{tr}(g)/2 \).

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
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 $\text{PSL}(2, \mathbb{R})$; this expresses the exceptional isomorphism of Lie groups $\text{PSL}(2, \mathbb{R}) = \text{SO}^+(2, 1)$. 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 $2 \times 2$ 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 $3 \times 3$ matrices

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

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

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

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 $\text{SL}(2, \mathbb{R}) \to \text{O}(2, 1)$ which factors through the quotient by the center $\pm \text{id}$, and realizes the isomorphism $\text{PSL}(2, \mathbb{R}) = \text{SO}^+(2, 1)$.

Another way to see this isomorphism is to think about the action of $\text{SL}(2, \mathbb{R})$ on the space of symmetric quadratic forms in two variables. A symmetric quadratic form $ax^2 + 2bxy + cy^2$ is represented by a symmetric $2 \times 2$ matrix $Q = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$ and the group $\text{PSL}(2, \mathbb{R})$ acts on such symmetric forms by $M \cdot Q = M^TQM$. 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 PSL$(2, \mathbb{R})$ action, since it is proportional to $\det(Q)$, and $\det(M) = \det(M^T) = 1$ for $M \in$ 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 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 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 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 $\theta$) where $\cosh(\ell/2) = \text{tr}(g)/2$.

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 $Q = \begin{pmatrix} a & \bar{z} \\ z & \bar{b} \end{pmatrix}$ 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 PSL$(2, \mathbb{C})$ acts on such forms by $M \cdot Q = M^TQM$ and preserves the discriminant $\det(QM) = \Delta := 4ab - z\bar{b} - \bar{z}a$ (which, again, is just proportional to $\det(Q)$), a nondegenerate form of signature $(3, 1)$. This exhibits the exceptional isomorphism PSL$(2, \mathbb{C}) = \text{SO}^+(3, 1)$.

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 $g \circ \varphi_t \circ \gamma$ 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 $\Gamma(\alpha \ast \gamma)(1) = \rho(\alpha)\Gamma(\gamma)(1)$ where $\ast$ 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 $\rho$ 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

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

**Proposition 2.2** (Thurston [13] 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 \( \sigma \) is open, this deformation gives rise to a deformation of the \((G, X)\) structure. \( \square \)

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 \( \phi_0 \) is a codimension 2 face in \( P_0 \), so it separates two adjacent codimension 1 faces \( \alpha_0 \) and \( \beta_0 \). Now, \( \beta_0 \) is glued to a face \( \alpha_1 \) in \( P_1 \), identifying \( \phi_0 \) with \( \phi_1 \), which separates \( \alpha_1 \) from \( \beta_1 \). Similarly, \( \beta_1 \) is glued to \( \alpha_2 \) in \( P_2 \), and we obtain a cycle of polyhedra \( P_i \) with codimension 2 faces \( \phi_i \) where each is glued to the next successively. Going once round the cycle takes \( \phi_0 \) 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 \( \phi_0 \) embeds isometrically in \( M \), and there is a hyperbolic structure on the interior of \( \phi_0 \) if and only if the dihedral angles in the \( P_i \) along the \( \phi_i \) add up to \( 2\pi \).

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 $\phi_i$ 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 of spherical polyhedra one dimension lower). This result is due to Poincaré, and is called:

**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 $\phi$ the dihedral angles at $\phi$ add up to $2\pi$ and the composition of the gluing isometries around $\phi$ are the identity on $\phi$. 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 [5].

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 $\phi$ 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.

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.

The edges of an ideal triangle are isometric to $\mathbb{R}$, so when ideal triangles are glued along a pair of edges, we must specify not just the orientation, but also the amount that the triangles are sheared. Each edge of an ideal triangle has a canonical midpoint (the foot of the perpendicular of the opposite vertex), so when we glue two edges we obtain a real-valued shear coordinate which measures the (signed) distance that each vertex is sheared to the right of the other. In the upper half-space model, we can fix the first triangle to have vertices $-1, 0, \infty$ and the second to have vertices $0, \infty, t$. Then the shear coordinate is $\log(t)$.

![Figure 4](image)

**Figure 4.** Five ideal triangles glued in a loop with hyperbolic holonomy around the vertex. The lifts of one of the triangles are in blue. The triangles accumulate on a “missing” geodesic (in red). The incomplete structure can be completed by adding the quotient of this missing geodesic, 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 $\log(t_i)$ (with indices taken cyclically), the holonomy of the developing map around the loop is given (in the upper half-plane model) by $z \rightarrow Tz + U$ where $T = \prod t_i$ and $U = t_1 + t_1t_2 + \cdots + T$. If $T = 1$, the holonomy transformation is parabolic, and the hyperbolic structure near the omitted vertex is complete. Otherwise, the holonomy is hyperbolic with translation length $\log(T)$, equal to the sum of the shear coordinates on the edges adjacent to the vertex. The hyperbolic structure can be completed to a surface with boundary by adding a geodesic of length $\log(T)$ which the ideal vertices “spin” around; see Figure 4.
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 \( S_3 \) 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 \( S_3 \), acting on simplex parameters by \( z \to 1/z \) and \( z \to 1/(1-z) \). In fact, 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

\[
z \to \frac{1}{1-z} \to \frac{z-1}{z} \to z
\]

We sometimes use the abbreviations \( z' := 1/(1-z) \) and \( z'' := (z-1)/z \). 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.

When two ideal tetrahedra are glued along faces, there is a unique isometry compatible with any identification of the vertices. If we glue finitely many simplices cyclically around an edge \( e \), we must check that we get an honest hyperbolic structure on \( e \). 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 \( z_i \), then the holonomy around \( e \) is given (in the upper half-space model) by the map \( z \to Tz \) where \( T = \prod z_i \). 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 \( e \), it is not enough that \( \prod z_i = 1 \) (this would just mean that the developing map has trivial holonomy around \( e \)) but the dihedral angles must add up to \( 2\pi \); i.e. \( \sum \log(z_i) = 2\pi i \). This is the edge equation associated to \( e \).

If \( M \) is obtained by gluing a finite set of ideal tetrahedra by isometric face pairings, then if the edge equations are all satisfied, one obtains a hyperbolic structure on \( M \). However, this hyperbolic structure might be incomplete. If \( \tilde{M} \) denotes the simplicial complex obtained by gluing honest simplices in the same combinatorial pattern, then \( M \) is homeomorphic to the complement of the vertices of \( \tilde{M} \). The link of each vertex \( v \) of \( \tilde{M} \) is a surface \( R_v \) which is made by gluing ideal vertex links. The link of an ideal vertex has a canonical Euclidean similarity structure, so the surfaces \( R_v \) come with developing maps \( D : \tilde{R}_v \to \mathbb{C} \) and holonomy representations \( \rho : \pi_1(R_v) \to \mathbb{C} \times \mathbb{C}^* \). Here the group \( \mathbb{C} \times \mathbb{C}^* \) acts on \( \mathbb{C} \) in the obvious way, with the first factor acting by translations, and the second by dilations. The hyperbolic structure on \( M \) is complete near \( v \) if the holonomy representation has trivial image in \( \mathbb{C}^* \). The cusp equations associated to a cusp say precisely that the dilations \( h(m) \), \( h(l) \) associated to holonomy around the meridian \( m \) and longitude \( l \) of the cusp are equal to 1. Each of the terms \( h(m) \) and \( h(l) \) is obtained as a product of cross-ratios of the ideal simplices meeting the cusp; thus they are each products of terms of the form \( z_i^{\pm 1} \) or \( \pm(1-z_i)^{\pm 1} \) where the \( z_i \) denote (marked) simplex parameters associated to the
ideal simplices. In conclusion, the edge and cusp equations (ignoring the condition on logarithms) can be expressed as integral algebraic equations in the simplex parameters, of quite a simple kind.

2.3.3. Edge and cusp equations. Suppose $M$ is obtained by gluing ideal tetrahedra with vertex links all tori. Suppose there are $t$ simplices and (after gluing) $e$ edges. Each simplex contributes 4 triangles to the vertex links, and 12 edges glued in pairs. Each edge contributes 2 vertices to the vertex links. Since the links are all tori, they have $\chi = 0$ 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 we can take a fundamental domain $P$ for the vertex link, and realize this as a (possibly non-convex) Euclidean similarity type of polygon made from triangles (the vertex links). Note that $P$ has an even number of edges, since the edges of $P$ are glued in pairs to make the cusp.

If $P$ is a Euclidean 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 Euclidean isometry $\phi_i$ taking $e_{i+1}$ to $e_i$ by an orientation-reversing isometry. The Euclidean similarity type determines the (complex) dilation $w_i$ of $\phi_i$. Since the composition of these isometries as we go once around $\partial P$ 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 the space of solutions of the edge equations has complex dimension equal to the number of cusps.

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 $\rho(m)$ is a (nontrivial) translation (i.e. the cusp equation holds for $m$), it follows that $\rho(l)$ must be too.

2.4. Hyperbolic Dehn surgery. Suppose $M$ is a 1-cusped hyperbolic 3-manifold, obtained by gluing (positively oriented) ideal simplices whose parameters solve the edge and cusp equations. The holonomy for each cusp $T$ is a representation $\rho : \pi_1(T) \to \mathbb{C}$, well-defined up to conjugacy, whose image is discrete and faithful. Nearby solutions of the edge equations parameterize incomplete structures on $M$, for which the representations $\rho : \pi_1(T) \to \mathbb{C} \rtimes \mathbb{C}^*$ have a nontrivial dilation part. We can conjugate such a representation into $\mathbb{C}^*$, acting on $\mathbb{C}$ by multiplication.

If $m, l$ are the meridian and longitude of $T$ with dilations $h(m), h(l)$, then we can choose branches $\log(h(m)) = \log(h(l)) = 0$ at the complete structures. As we deform the complete structures, both $\log(h(m))$ and $\log(h(l))$ become nonzero, and (generically), there are unique real numbers $p, q$ for which

$$p \log(h(m)) + q \log(h(l)) = 2\pi i$$

For typical $p, q$ the representation $\rho$ is indiscrete. 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 cyclic group. The quotient of \( \tilde{\gamma} \) by \( \langle \rho(m), \rho(l) \rangle \) is a closed geodesic \( \gamma \) which completes \( M \), giving rise to a hyperbolic structure on the closed manifold \( M_{p/q} \) obtained by doing Dehn filling on \( M \) along the slope \( p/q \).

With this background, we can now prove

**Theorem 2.4** (Thurston’s hyperbolic Dehn surgery Theorem [13], 5.8.2). Let \( M \) be a 1-cusped hyperbolic 3-manifold with torus cusp \( T \) and coordinates \( m, l \). The dilation \( h(m) \) holomorphically parameterizes the space of (not necessarily complete) hyperbolic structure on \( M \) near the complete structure.

Moreover, for all but finitely many \( (p, q) \) there is a deformation of the hyperbolic structure with \( p \log(h(m)) + q \log(h(l)) = 2\pi i \) which can be completed to a closed manifold homeomorphic to \( M_{p/q} \). The manifolds \( M_{p/q} \) converge geometrically on compact subsets (after choosing suitable basepoints) to \( M \) as \( (p, q) \to \infty \).

**Proof.** We assume for simplicity that the complete manifold \( M \) admits an ideal triangulation with all simplices positively oriented as in § 2.3.2, although this is not strictly necessary. Solutions to the edge equations near the complete structure give rise to incomplete structures, and all nearby incomplete structures are of this form. By Proposition 2.2 some neighborhood of the complete structure may be identified with an open subset of the space \( X \) of conjugacy classes of representations \( \pi_1(M) \to \text{PSL}(2, \mathbb{C}) \) containing the class of the discrete faithful representation \( \rho_0 \) corresponding to the complete structure. In this way we may think of \( X \) (at least locally) as a complex analytic variety with the \( z_i \) as holomorphic coordinates. We have proved by our dimension count that \( X \) has dimension at least 1 near \( \rho \). This is the only place in the argument where an ideal triangulation is
used; in fact the argument works just as well if some simplices are degenerate (i.e. have real parameter) at the complete structure, and this can always be achieved.

The traces \( \text{tr}(\rho(m)) \) and \( \text{tr}(\rho(l)) \) are holomorphic functions on \( X \) and take the value 2 near \( \rho_0 \) (for a suitable lift to \( \text{SL}(2, \mathbb{C}) \)). Since \( \rho(m) \) and \( \rho(l) \) commute, one is parabolic if the other is. A deformation of the complete structure keeping both parabolic will stay complete, and the complete structure is unique by Mostow-Prasad Rigidity, as we will see in § 3.1.5; so the trace map \( X \to \mathbb{C}^2 \) has 1-dimensional image near \( \rho \).

Suppose we have chosen coordinates such that for the complete structure, 

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

where \( c \) is a complex number with positive imaginary part. At a nearby \( \rho \) the matrices \( \rho(m) \) and \( \rho(l) \) are hyperbolic and commute, so they have common eigenvectors \( \begin{pmatrix} 1 \\ 1 \end{pmatrix} \) and \( \begin{pmatrix} 1 \\ -1 \end{pmatrix} \). Suppose the eigenvalues of \( \rho(m) \) and \( \rho(l) \) are \( \mu^\pm \) and \( \lambda^\pm \) respectively, where \( \mu, \lambda \) are the eigenvalues for the first eigenvector. By convention we have \( h(m) = \mu \) and \( h(l) = \lambda \). Following Thurston, we compute

\[
\begin{pmatrix} 1 \\ 1 \end{pmatrix} \sim \rho(m) \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \frac{1}{\epsilon_1 - \epsilon_2} \rho(m) \left( \begin{pmatrix} 1 \\ \epsilon_1 \end{pmatrix} - \begin{pmatrix} 1 \\ \epsilon_2 \end{pmatrix} \right) = \frac{1}{\epsilon_1 - \epsilon_2} \left( \frac{\mu - \mu^{-1}}{\mu \epsilon_1 - \mu^{-1} \epsilon_2} \right)
\]

and therefore \( \mu - \mu^{-1} \sim \epsilon_1 - \epsilon_2 \), and similarly (using \( \rho(l) \) in place of \( \rho(m) \) above), \( \lambda - \lambda^{-1} \sim c(\epsilon_1 - \epsilon_2) \). Thus for \( \mu, \lambda \) close to 1, 

\[
\frac{\log \lambda}{\log \mu} \sim \frac{\lambda - 1}{\mu - 1} \sim \frac{\lambda - \lambda^{-1}}{\mu - \mu^{-1}} \sim c
\]

and thus the generalized coordinates \( p, q \) almost satisfy

\[
\begin{align*}
(2.3) \quad p + qc & \sim \frac{2\pi i}{\log \mu} \\
\end{align*}
\]

Since \( \mu \) takes values in a neighborhood of 1 for \( \rho \) in a neighborhood of \( \rho_0 \), it follows that all but a compact set of \( p, q \) are realized near \( \rho_0 \). For integer \( p, q \) we obtain hyperbolic structures on the closed manifolds \( M_{p/q} \). By Mostow Rigidity (see Theorem 3.1), such structures are unique up to isometry, and therefore the map from \( X \) to \( \mu \) is locally injective near \( \rho_0 \). The proof follows. \( \square \)

2.4.1. Generalized Dehn surgery. When \( p, q \) are integers for which

\[
(2.4) \quad p \log(h(m)) + q \log(h(l)) = \pm 2\pi i
\]

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.
2.5. **Examples of hyperbolic 3-manifolds.**

*Example 2.5 (Doubling).* Let $P$ be a compact 3-dimensional hyperbolic polyhedron with totally geodesic faces, and all dihedral angles of the form $\pi/n$ 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 $\Delta$ by each of these four planes to obtain a truncated tetrahedron with dihedral angles all equal to $\alpha$ and $\pi/2$. Taking $\alpha = \pi/n$ for $n > 3$ we obtain infinitely many (incommensurable) examples this way.

*Example 2.6 (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 6.

![Figure 6](image)

*Figure 6.* Two regular ideal simplices glued with this pairing gives a complete hyperbolic manifold homeomorphic to the complement of the figure 8 knot in $S^3$.

Six simplices meet (locally) along each edge, and because the simplex parameters are all equal to $e^{2\pi i/6}$ the edge equations are satisfied. The fundamental domain for the cusp is a parallelogram formed from 8 equilateral triangles. The holonomy is parabolic, and the structure is complete.

*Example 2.7 (Alternating link complements).* It is a demanding exercise in visualization to translate a knot or link projection into a combinatorial ideal triangulation of the complement, but there is a systematic method which works well for alternating links.

Suppose $L$ is a link projection. We can embed $L$ in a graph $\Gamma$ by adding one “vertical” edge for each crossing, which joins the overcrossing point to the undercrossing point. Complementary regions to the projection are polygons, whose edges are arcs of $L$ joining adjacent crossings. A complementary $n$-gon $P$ to the projection determines a $2n$-gon $\tilde{P}$
obtained by inserting a vertical edge at each vertex of $P$. Then we can obtain an ideal $n$-gon $P'$ from $\bar{P}$ by removing the original edges of $P$ (which lie on $L$) from $\bar{P}$, replacing them by ideal vertices. Thus: the edges of $P'$ correspond to the crossings on the boundary of the region $P$.

Now let’s suppose $L$ is alternating. The complement $S^3 - L$ is obtained by gluing two (combinatorial) ideal polyhedra $B^\pm$ defined as follows. Each of $\partial B^\pm$ has one copy of each ideal polygon $P'$ as a face, and all faces arise this way. We glue $B^+$ to $B^-$ along their boundaries by gluing each $P'$ in $\partial B^+$ to the $P'$ in $\partial B^-$ by the “identity” map. It remains to describe how the copies of $P'$ fit together combinatorially in $\partial B^+$ and in $\partial B^-$. Suppose $P, Q$ are complementary polygons to $L$ which share an edge $e \subset L$ oriented to run from an undercrossing $e^-$ to an overcrossing $e^+$. Note that the crossings $e^\pm$ will correspond to pairs of edges of $P^+$ and $Q^+$ in $\partial B^+$ and in $\partial B^-$. Suppose with respect to the orientation on $e$ that $P$ is on the left and $Q$ is on the right. Then the copies of $P'$ and $Q'$ share one edge in $\partial B^+$ and one edge in $\partial B^-$ as follows:

- in $\partial B^+$, $P'$ and $Q'$ meet along $e^-$; and
- in $\partial B^-$, $P'$ and $Q'$ meet along $e^+$.

This determines the way the different $P'$ meet in $\partial B^+$ and in $\partial B^-$, and thus the combinatorics of the gluing.

If some complementary regions to $L$ are bigons, they give rise to a pair of bigons in $\partial B^+$ and $\partial B^-$ which may be collapsed to edges without changing the topology of the quotient. This simplification is useful in practice, since an honest geodesic ideal polyhedron can’t have faces which are bigons.

The Figure 8 knot $K$ has a projection with 6 complementary regions consisting of 4 triangles and 2 bigons. After collapsing bigons, we obtain $S^3 - K$ by gluing two ideal tetrahedra, as in Example 2.6. The Borromean rings $L$ in its standard projection has 8 complementary triangle regions. By symmetry, both $B^\pm$ in this case are (ideal) octahedra, and $S^3 - L$ can be realized geometrically by gluing two regular ideal octahedra in a suitable combinatorial pattern.

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 $\text{Out}(\pi_1(M))$ 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 \( d(\tilde{g}\tilde{f}(p), p) \leq C \) 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

\[
\frac{1}{K}d_{\tilde{N}}(\tilde{f}(p), \tilde{f}(q)) - K \leq d_{\tilde{M}}(p, q) \leq Kd_{\tilde{N}}(\tilde{f}(p), \tilde{f}(q)) + K
\]

If \( \gamma \) is a geodesic in \( \mathbb{H}^n \), we can define a function \( \rho : \mathbb{H}^n \to \mathbb{R}^+ \) to be the distance to \( \gamma \). 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 \( \gamma \) the image \( \tilde{f}(\gamma) \) is contained within distance \( O(\log(K)) \) of some unique geodesic \( \delta \), and the map \( \tilde{f} \) extends continuously (by taking endpoints of \( \gamma \) to endpoints of \( \delta \) as above) to a homeomorphism \( \tilde{f}_\infty : S_{\infty}^{n-1} \to S_{\infty}^{n-1} \). 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 \( \alpha \), denoted \( \| \alpha \| \), is the infimum of \( \| z \| \) over all real \( k \)-cycles \( z \) representing \( \alpha \).

The name Gromov “norm” is misleading, since it could easily be 0 on some nonzero \( \alpha \). 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 \( s : C_\ast(M; \mathbb{R}) \to C_\ast(M; \mathbb{R}) \) 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 \( \sigma \) 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 \( \text{id} \) 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 \( \|\| [M] \| \leq \text{volume}(M)/v_n \), 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 \( \Delta \) 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 \( \epsilon \) sufficiently small, any other totally geodesic simplex \( \Delta' \) whose vertices are obtained from those of \( \Delta \) 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 \( \Delta \) to \( \mathbb{H}^n \), and we can put an invariant locally finite measure \( \mu \) 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 \( \mu \) as a “measurable” singular \( n \)-chain in \( M \), where by convention we parameterize each \( \Delta \) 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 \( \phi \) of \( \Delta \) there is another isometric map \( \Delta \to M \) obtained by reflection in \( \phi \), and the contributions of these two maps to \( \phi \) 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 \( \Delta \).

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 \( pg := gp_1 \). Now, if we denote an \((n+1)\)-tuple \((g_0, \ldots, g_n) \in \pi_1(M)^{n+1} \) by \( \vec{g} \) for short, we define \( c(\vec{g}) \) to be the \( \mu \)-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 \( pg_i \). 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. \( \Box \)
It is a theorem of Haagerup and Munkholm [7] 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 $\Delta$ 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 $S_{\infty}^{n-1}$ 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 $n \geq 3$ 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 $\Gamma$ on $S_{\infty}^{n-1}$ generated by reflections in the side of a regular ideal simplex. Orbits of $\Gamma$ on $S_{\infty}^{n-1}$ 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 $\text{Isom}(\mathbb{H}^n)$ 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 $\text{volume}(M) \geq d \cdot \text{volume}(N)$, 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 $\text{volume}(M) > d \cdot \text{volume}(N)$ unless $f$ is homotopic to a covering map of degree $d$.

Since $f : M \to N$ is not a priori $\pi_1$-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 $\tilde{f}_\infty : S_{\infty}^{n-1} \to S_{\infty}^{n-1}$. 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 $g_0 = \text{id}$, 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 § 6.4) 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 $\tilde{f}_\infty : S_{\infty}^{n-1} \to S_{\infty}^{n-1}$ 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. If $f : M \to N$ is a homotopy equivalence between complete finite volume hyperbolic manifolds, the Mostow-Prasad rigidity theorem says that $f$ is homotopic to an isometry. This can be proved along similar lines to the
arguments above. A 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 [8]. The notation \( M_{(0,\epsilon]} \) for the \( \epsilon \)-thin part of \( M \) (where the injectivity radius is at most \( \epsilon \)) is explained in § 3.2. For each big \( \ell \) we can find small constants \( 0 < \epsilon < \epsilon_1 \) where \( M_{(0,\epsilon]} \) 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 \( v_\epsilon \), and with all edges of length close to \( \ell \), 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 \( \partial z \) is within \( \ell \) 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 \( H_n(M, M_{(0,\epsilon]} \cong H_n(\bar{M}, \partial \bar{M}) \) where the latter map is induced by a deformation retraction, which induces a chain map of norm 1.

The rest of the proof of Mostow-Prasad rigidity is the same.

3.2. Margulis lemma. Let \( M \) be a complete hyperbolic \( n \)-manifold (not necessarily compact). For any \( \epsilon > 0 \) we define the \( \epsilon \)-thin part of \( M \), denoted \( M_{(0,\epsilon]} \), to be the closed subset where the injectivity radius is at most \( \epsilon^2 \), and the \( \epsilon \)-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 \( \epsilon_n \) so that the \( \epsilon_n \)-thin part of any complete hyperbolic \( n \)-manifold has a very simple topology. Explicitly:

**Theorem 3.4** (Margulis Lemma). In each dimension \( n \) there is a positive constant \( \epsilon_n \) 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 \( \epsilon \) so that if \( |g|, |h| < \epsilon \), we have an inequality

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

From this we deduce the following lemma:

**Lemma 3.5.** For any Lie group \( G \) with a left-invariant metric there is an \( \epsilon \) so that if \( \Gamma \) is a discrete subgroup of \( G \), and \( \Gamma_{\epsilon} \) is the subgroup of \( \Gamma \) generated by elements \( g \) with \( |g| < \epsilon \), then \( \Gamma_{\epsilon} \) is nilpotent.
Proof. Because of the identity \([a,bc] = [a,b][b,[a,c]]\) (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 \(g_0, \ldots, g_m \in \Gamma\) have \(|g_i| < \epsilon\) then

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

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

Note that whereas \(\epsilon\) depends only on \(G\), the nilpotence depth \(m\) of \(\Gamma_\epsilon\) may depend on \(\Gamma\).

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 \(\text{Isom}(\mathbb{H}^n)\).

Define a metric on \(\text{Isom}(\mathbb{H}^n)\) by \(|g| = d(\tilde{p}, gp\tilde{p}) + |\tau(g)|\) where \(\tau(g) \in O(n)\) is the rotation of \(T_p\mathbb{H}^n\) induced by applying \(g\) at \(\tilde{p}\) and then parallel transporting back to \(\tilde{p}\) along the geodesic from \(\tilde{p}\) to \(\tilde{p}\), and \(|\cdot|\) is some bi-invariant metric on \(O(n)\). We claim that for any \(\epsilon\) there is an \(\epsilon'\) so that if \(\Gamma'\) is the subgroup of \(\pi_1(M)\) generated by \(g\) with \(d(\tilde{p}, gp\tilde{p}) \leq \epsilon'\), the group \(\Gamma'\) contains with finite index \(\Gamma\), the subgroup of \(\pi_1(M)\) generated by elements with \(|g| < \epsilon\). Choosing \(\epsilon\) as in Lemma 3.5 and taking \(\epsilon_n = \epsilon'\) the proof of the first part of Theorem 3.4 will be complete. Let \(S'\) denote the set of \(g \in \pi_1(M)\) with \(d(\tilde{p}, gp\tilde{p}) < \epsilon'\) 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 \(\Gamma'\) as a product

\[
w = g_1g_2g_3 \cdots g_m
\]

where each \(g_i \in S'\). Now, it is not quite true that \(\tau\) is a homomorphism from \(G\) to \(O(n)\), but the difference between \(\tau(gh)\) and \(\tau(g)\tau(h)\) is controlled by the curvature tensor, which is quadratic in \(d(\tilde{p}, gp\tilde{p})\) and \(d(\tilde{p}, h\tilde{p})\). Since \(O(n)\) is compact, there is a \(C\) depending only on \(\epsilon\) such that for any \(C\) elements of \(O(n)\) there are two with distance at most \(\epsilon/8\). Thus we may find distinct indices \(i, j \leq C\) (assuming \(m \geq C\)) so that \(\tau(g_{i+1} \cdots g_j) < \epsilon/4\). We may furthermore assume that \(\epsilon' < \epsilon/2\) so that the product \(g\) of at most \(C\) elements of \(S'\) has \(d(\tilde{p}, gp\tilde{p}) < \epsilon'\), and thus \(|g_{i+1} \cdots g_j| < \epsilon/2\).

Now, the metric on \(\text{Isom}(\mathbb{H}^n)\) is not conjugation invariant, but it is invariant under conjugation by \(O(n)\). Since \(O(n)\) is compact, so we may suppose that the metric on \(\text{Isom}(\mathbb{H}^n)\) has the property that \(|g^h| < 2|g|\) for \(|g| < \epsilon\) and for \(h\) sufficiently close to \(O(n)\); i.e. (taking \(\epsilon'\) small enough) for arbitrary \(h\) with \(d(\tilde{p}, h\tilde{p}) < C\epsilon'\).

Thus we may rewrite

\[
g_1g_2 \cdots g_j = (g_{i+1} \cdots g_j)^{g_{i+1} \cdots g_j}g_i \cdots g_i
\]

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 \(\Gamma\) has finite index in \(\Gamma'\) 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 \(\text{Isom}(\mathbb{H}^n)\). Consider some component \(K\) of \(M_{[0,\epsilon]}\) with fundamental group \(\Gamma\). 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 \(\Gamma\) contains a hyperbolic element \(g\) with fixed points \(p^\pm\) then every element of \(\Gamma\) must fix both \(p^\pm\). Since \(\Gamma\) is torsion-free and discrete, it follows that \(\Gamma = \mathbb{Z}\) in this case, and \(K\) is a tube around an embedded geodesic.
If $\Gamma$ 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 $\Sigma$ 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 3.6** (Gauss-Bonnet). Let $\Sigma$ be a closed Riemannian 2-manifold. Then

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

where $K$ denotes the Gauss curvature.

**Proof.** Sectional curvature is tensorial, and therefore on a surface is captured by a 2-form $K \text{d} \text{area}$ which measures the amount of rotation of the tangent space under parallel transport around an infinitesimal parallelogram. By integrating this relationship we see that $\int_{\Omega} K \text{d} \text{area}$ is equal (up to integer multiples of $2\pi$) to the rotation of the tangent space under parallel transport around the oriented boundary $\partial \Omega$ for any domain $\Omega$. Taking $\Omega = \Sigma$ we see that $\int K \text{d} \text{area}$ is an integer multiple of $2\pi$, 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 $2\pi$, and using Euler’s formula $\chi = F - E + V$ the theorem follows.  

Even if the surface $\Sigma$ is not smooth everywhere, providing parallel transport makes sense on “enough” curves, it is possible to define curvature as a (signed) Radon measure on $\Sigma$ in such a way that the Gauss-Bonnet theorem is still valid. For example, if $\Sigma$ 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

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

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

\[(3.3) \quad D(z) - D(w) + D\left(\frac{w}{z}\right) - D\left(\frac{1-w}{1-z}\right) + D\left(\frac{1-w^{-1}}{1-z^{-1}}\right) = 0\]

It turns out that \(D\) as above is the *Bloch-Wigner dilogarithm*, defined by

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

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 Schlafli formula, will be given in the proof of Proposition 3.10.

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

\[(3.5) \quad \Lambda(\theta) := -\int_0^\theta \log|2\sin t|dt\]

The volume of an ideal simplex has a very elegant description in terms of \(\Lambda\):

**Proposition 3.7.** If \(\Delta\) is an ideal simplex with dihedral angles \(\alpha, \beta, \gamma\) 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 \(\infty\). 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 \(\alpha, \beta, \gamma\) (in pairs). So it suffices to compute the volume of the region \(\sigma_\alpha\) above one of these six triangles, say with angle \(\alpha\), and show it is \(\Lambda(\alpha)/2\).

We compute

\[
\text{volume}(\sigma_\alpha) = \int_0^{\cos \alpha} dx \int_0^{x \tan \alpha} dy \int_\frac{\sqrt{1-x^2-y^2}}{z^3} dz
\]

\[(3.6) \quad = \frac{1}{2} \int_0^{\cos \alpha} dx \int_0^{x \tan \alpha} dy \frac{dy}{1-x^2-y^2}
\]

\[= \frac{1}{4} \int_0^{\cos \alpha} \log \frac{\sqrt{1-x^2} \cos \alpha + x \sin \alpha}{\sqrt{1-x^2} \cos \alpha - x \sin \alpha \sqrt{1-x^2}} dx\]
Doing the substitution $x = \cos t$ gives
\[
\text{volume}(\sigma_\alpha) = -\frac{1}{4} \int_{\pi/2}^\alpha \log \frac{\sin t \cos \alpha + \cos t \sin \alpha}{\sin t \cos \alpha - \cos t \sin \alpha} \, dt
\]
\[
= \frac{1}{4} \int_{\alpha}^{\pi/2} \log \frac{\sin t + \alpha}{\sin t - \alpha} \, dt
\]
\[
= \frac{1}{4} \int_{2\alpha}^{\alpha+\pi/2} \log |2 \sin t| \, dt - \frac{1}{4} \int_0^{\pi/2-\alpha} \log |2 \sin t| \, dt
\]
\[
= \frac{1}{4} \left( -\Lambda(\alpha + \pi/2) + \Lambda(2\alpha) + \Lambda(\pi/2 - \alpha) \right)
\]
(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 $\theta$. Taking $\theta = \pi/2$ gives $2\Lambda(\pi) = \Lambda(\pi)$ so $\Lambda(\pi) = 0$ and we see that $\Lambda$ is periodic with period $\pi$. Since it is evidently odd (because the integrand is even), it follows that $\Lambda(\pi/2) = 0$. Oddness and $\pi$-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. \(\square\)

3.3.3. Schlafli’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 $\text{volume}(t)$ denote the $n$-dimensional volume of $P(t)$. Then there is a remarkable differential formula for the variation of $\text{volume}(t)$ as a function of $t$ due essentially to Schlafli:

**Theorem 3.8 (Schlafli’s formula).** With notation as above, there is a differential identity
\[
\frac{d\text{volume}(t)}{dt} = -\frac{1}{n-1} \sum_e \ell_e(t) \frac{d\theta_e(t)}{dt}
\]
(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 $\pi$ by a parabolic motion fixing a point of $\pi$ at infinity, and with (horocircular) orbits perpendicular to $\pi$. 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 $\pi$ is vertical and parallel to the $y$-$z$ plane, and intersects $P$ in a triangle $\Delta$. Cyclically label the oriented edges of $\Delta$ as $e_1, e_2, e_3$ so that $e_1$ is contained in the intersection of $\pi$ with the unit hemisphere centered at the origin (in the $x$-$y$ plane). The $x$ coordinate is constant on $\Delta$, and we consider a deformation of $P$ obtained by moving $\Delta$ by translating it by $dx$. For this motion, there is a formula
\[
\frac{d\text{volume}}{dx} = \int_{\Delta} dy \wedge dz = \frac{1}{2} \int_{\partial\Delta} \frac{dy}{z^2}
\]
(3.9)
where the second equality follows by Stokes’ theorem.

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

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

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. \hfill \Box

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 $\text{volume}(P(t)) = \text{volume}(M(t))$. 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 $2\pi$. Thus Schläfli immediately shows that $\text{volume}(M(t))$ 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 § 2.4.1, 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 $\text{volume}(M_{p/q}) < \text{volume}(M)$.

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 $\omega$, whose limit (the supremum, which is not achieved) is equal to $\text{volume}(M)$.

In fact, it is possible to estimate the difference in volumes $\text{volume}(M) - \text{volume}(M_{p/q})$ to leading order (for big $p, q$) in terms of the cusp shape $c := a + bi$. Combining Equation 2.4 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

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

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

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

and therefore, by using Schläfli and integrating, we get the following estimate, first obtained by Neumann-Zagier [11] using direct methods:
Proposition 3.9. 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

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

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 3.10. If \( \Delta(z) \) is an ideal simplex with parameter \( z \) then

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

Proof. In the upper half-space put the four vertices of \( \Delta \) at 0, 1, \( z \), \( \infty \). Let \( H_\infty \) 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 \( \Delta \). 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 \( \infty 0, \infty 1, \infty z, 01 \) 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 \( 2\pi \), 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 \( z \):

\[
(3.15) \quad \frac{d\text{volume}(\Delta(z))}{dz} = \log |z| \frac{d\arg(1 - z)}{dz} - \log |1 - z| \frac{d\arg z}{dz}
\]

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

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

and the two terms under the integral sum to \( \text{Im} (\log(1 - z) d(\log z)/dz) \), completing the proof. \( \square \)
3.3.4. **Thurston-Jorgenson theorem.** The following remarkable theorem follows almost formally from what we have done so far:

**Theorem 3.11** (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 $\omega^\omega$.

**Proof.** From Thurston’s hyperbolic Dehn surgery Theorem 2.4, Mostow’s Rigidity Theorem 3.1 (and its strengthening due to Prasad), and the Neumann-Zagier volume formula in Proposition 3.9, the theorem will follow once we show that for any positive number $V$, there is a finite set of finite volume cusped manifolds $M_1, \ldots, M_n$ (where $n$ and the $M_j$ depend on $V$ of course) such that every finite volume complete hyperbolic manifold $M$ with $\text{volume}(M) \leq V$ 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 3.4. If $V$ is fixed, there are only finitely many possibilities for the topology of the thick part $M_{(\epsilon, \infty)}$ for any $M$ with $\text{volume}(M) \leq V$. 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 3.9 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 $\kappa$ is an infinite ordinal which is a “polynomial” in $\omega$; 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 $\text{volume}(M)$ where $M$ is finite volume but noncompact with $j$ cusps is a limit volume; i.e. it is of the form $v_\kappa$ for $\kappa = a_k\omega^k + \cdots + a_n\omega^n$ with $a_k$ nonzero, for some $k \geq j$.

**Example 3.12** (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 $v_\kappa^+$ associated to “simple” ordinals $\kappa$ 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, [6]);
- $v_{\omega}^+ \sim 2.02988 \cdots$ 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, [3]); 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 \( v^+_\omega \) is twice the volume of the regular ideal simplex, and \( v^+_\omega \) is the volume of the regular ideal octahedron, and in fact the associated minimal volume manifolds can be obtained by gluing up these polyhedra.

4. Quasiconformal deformations and Teichmüller theory

4.1. Complex analysis.

4.1.1. Equicontinuity and the Schwarz lemma. The Cauchy-Riemann equations let one control higher derivatives in terms of lower ones. This lets us prove equicontinuity and (hence) precompactness for families under very mild hypotheses. The key property of the Cauchy-Riemann equations is ellipticity, which manifests itself in various versions of the maximum modulus principle. One of the most useful corollaries of the maximum modulus principle is the Schwarz Lemma:

**Theorem 4.1** (Schwarz Lemma). Let \( f \) be a holomorphic map taking the (open) unit disk \( D \) inside itself, normalized to have \( f(0) = 0 \). Then \( |f(z)| \leq |z| \) for all \( z \in D \) with equality for some \( z \neq 0 \) if and only if \( f \) is of the form \( f(z) = e^{i\theta}z \) for some real \( \theta \).

**Proof.** Since \( f(0) = 0 \), write \( f(z) = zg(z) \) for some \( g \) analytic on \( D \). Then \( |g(z)| \leq |f(z)/z| \leq 1/r \) on the circle \( |z| = r \), and therefore by the maximum modulus principle, \( |g(z)| \leq 1/r \) for \( |z| \leq r \). Letting \( r \to 1 \) we obtain \( |g(z)| \leq 1 \) everywhere, and again by the maximum modulus principle there is equality somewhere in the interior iff \( g(z) = e^{i\theta} \) for some constant \( \theta \). \qed

In particular, \( |f'(0)| \leq 1 \) with equality iff \( f(z) = e^{i\theta}z \). Conjugating such an \( f \) by an element of \( \text{PSU}(1,1) \) we see that bounded holomorphic functions in a compact domain have bounded derivatives, and hence (by induction) derivatives of all orders are uniformly bounded. From this follows Montel’s theorem:

**Theorem 4.2** (Montel’s Theorem). Let \( f_n \) be a family of locally bounded holomorphic functions on a domain \( \Omega \). Then \( f_n \) has a subsequence which converges uniformly on compact subsets of \( \Omega \) to some limit \( f \) (which is necessarily holomorphic).

4.1.2. Riemann mapping theorem. From the Schwarz Lemma and Montel’s Theorem one can deduce the Riemann mapping theorem:

**Theorem 4.3** (Riemann Mapping Theorem). Let \( \Omega \) be a simply-connected open subset of the Riemann sphere \( \hat{\mathbb{C}} \). Then \( \Omega \) is holomorphically equivalent to exactly one of \( \hat{\mathbb{C}} \), \( \mathbb{C} \) or \( D \).

**Proof.** It suffices to treat the case that \( \Omega \subset \mathbb{C} \) omits at least 2 points (since \( \mathbb{C}^* \) is not simply-connected). By a Möbius transformation we can put these two points at 0 and 1728 and then use the inverse of Klein’s \( j \)-function to uniformize the universal cover of \( \mathbb{C} - \{0,1728\} \) as the upper half-plane, which is conformally equivalent to the disk; the conclusion is that we can find some holomorphic embedding \( h : \Omega \to D \), which can be normalized to send some fixed \( p \in \Omega \) to 0.

Now let \( \mathcal{F} \) denote the set of all holomorphic embeddings \( \Omega \to D \) sending \( p \) to 0 together with the constant map \( \Omega \to p \). By Montel’s theorem \( \mathcal{F} \) is compact, so we may find \( f \in \mathcal{F} \)
maximizing $|f'(p)|$. We claim that $f$ is onto; i.e. that $f : \Omega \to D$ is a holomorphic isomorphism. If not, then let $w \in D$ be an omitted value, and define

$$F(z) = \sqrt{\frac{f(z) - w}{1 - \bar{w}f(z)}}$$

by taking some branch of the square root (which we can do because $f(\Omega)$ is simply-connected and avoids $w$) and observe that $F$ is single-valued and takes values in $D$. Composing $F$ with a suitable element of $\text{PSU}(1,1)$ produces

$$G(z) = \frac{F(z) - F(p)}{1 - F(p)F(z)}$$

which sends $p$ to 0, and satisfies

$$|G'(p)| = \frac{|F'(p)|}{1 - |F(p)|^2} = \frac{1 + |w|}{2\sqrt{|w|}}|f'(p)| > |f'(p)|$$

contrary to the definition of $f$.

4.1.3. Nonlinearity, Schwarzian derivative. It is an important and useful fact that even when the uniformizing map $f$ promised by the Riemann mapping theorem is hard to find explicitly, it is sometimes easier to find a differential equation satisfied by $f$ (this is the same phenomenon that we encountered already in § 1.1).

The group of holomorphic automorphisms of $\hat{C}$ is generated by translations ($z \to z + c$), dilations ($z \to cz$), and inversions ($z \to 1/z$). We describe differential operators invariant under progressively more of these symmetries.

First, taking derivatives itself is invariant under translation. Thus, the transformation $f \to f + c$ preserves $f'$. Second, dilations can be transformed to translations after taking logs. Thus $f \to cf$ preserves $(\log(f))'$. Taking these two things together motivates the definition of the Nonlinearity

$$N(f) := (\log(f'))' = f''/f'$$

The function $f$ can be recovered from $N(f)$ by integrating twice: $f(z) = \int e^{\int N(f)}$ up to the ambiguity of composition with an arbitrary dilation and translation.

Suppose $P$ is a (Euclidean planar) polygon with angles $\alpha_i$ at vertices $p_i$ in cyclic order, and suppose $f : H \cup \infty \to P$ uniformizes $P$ by the upper half-plane plus infinity. If $q_i \in \mathbb{R}$ is the preimage of $p_i$, then near $q_i$ the map $f$ looks like $z \to z^{\alpha_i/\pi}$, up to composition with a translation and a dilation. The transcendental function $g : z \to z^\alpha$ satisfies $N(g) = (\alpha - 1)/z$ which is algebraic. Thus the nonlinearity of the uniformizing map $f$ satisfies $N(f) = (\alpha_i - \pi)/z\pi + O(1)$ at $q_i$. Note further that $f$ (and therefore $N(f)$) is nonsingular throughout the rest of $H \cup \infty$.

By the Schwarz reflection principle we can analytically continue $f$ across each of the intervals $(q_i, q_{i+1})$ and the different results map the lower half-plane to the polygons obtained by reflecting $P$ in its sides. Continuing inductively, $f$ can be analytically continued to a single-valued holomorphic map from the universal cover of $\hat{C} - \cup_i q_i$. In general, $f$ will not be single-valued on $\hat{C}$, but we have just seen that its nonlinearity $N(f)$ is; it is a meromorphic function with first order poles exactly at the points $q_i$. In particular, $N(f)$ is an
algebraic function, and since it can’t have a pole at infinity (unless some \( q_i \) is at infinity), it is necessarily of the form \( N(f) = C + 1/z (\sum(\alpha_i/\pi) - 1) \) for a suitable constant \( C \). We therefore obtain the explicit integral formula for so-called Schwarz-Christoffel mapping:

**Proposition 4.4** (Schwarz-Christoffel). Let \( f : H \cup \infty \to P \) uniformize a Euclidean polygon with angles \( \alpha_i \) at points \( p_i \) with preimages \( q_i \in \mathbb{R} \). Then for suitable constants \( K_1, K_2 \)

\[
(4.5) \quad f(\zeta) = K_1 + \int_0^{\zeta} \frac{K_2}{(z - q_1)^{1-(\alpha_1/\pi)}(z - q_2)^{1-(\alpha_2/\pi)} \ldots} dz
\]

Thus in principle, determining \( f \) comes down to figuring out the location of the \( q_i \). For small numbers of vertices (e.g. if \( P \) is a triangle) we can precompose by an automorphism of \( H \) so that three of the \( q_i \) are at 0, 1, \( \infty \), making the rest easier to find.

The transformation \( f \to 1/f \) takes \( N(f) \) to \( N(1/f) = N(f) - 2f''/f' \). From this one may check that the following operator

\[
(4.6) \quad S(f) := N'(f) - N(f)^2/2 = \frac{f'''}{f'} - \frac{3}{2} \left( \frac{f''}{f'} \right)^2
\]

known as the Schwarzian derivative of \( f \), is completely invariant under Möbius transformations.

If \( P \) is a spherical polygon in \( \hat{C} \) whose arcs are round circles, the map \( f : H \cup \infty \to P \) can be composed near each vertex \( p_i \) with a Möbius transformation in such a way that its nonlinearity has a simple pole there, and therefore we see that \( S(f) \) has second order poles at each of the preimages \( q_i \) and by the same reasoning as above, we deduce that \( S(f) \) has a single-valued meromorphic extension to all of \( \hat{C} \) which is rational, of the form

\[
(4.7) \quad S(f) = c + \sum_i a_i/(z - q_i)^2 + b_i/(z - q_i)
\]

for suitable constants \( a_i, b_i, c \) (which are unfortunately rather hard to determine explicitly). Knowing the Schwarzian, the Nonlinearity can be recovered by solving the Ricatti equation

\[
(4.8) \quad N' - N^2/2 - S = 0
\]

and then knowing \( N \) we can recover \( f \).

4.2. Riemann surfaces. A Riemann surface \( S \) is just a smooth surface with an atlas of charts modeled on \( \mathbb{C} \) with holomorphic transition functions. Since the definition is local, any covering space of a Riemann surface is a Riemann surface.

Beyond the Riemann mapping theorem there is the uniformization theorem, first proved by Poincaré in 1907:

**Theorem 4.5** (Uniformization Theorem). Let \( S \) be any Riemann surface. Then the universal cover \( \tilde{S} \) is holomorphically equivalent to exactly one of \( \hat{C}, \mathbb{C} \) or \( D \).

**Proof.** The uniqueness is obvious. Triangulate \( \tilde{S} \) by a locally finite triangulation with real analytic edges for which each triangle is contained in a holomorphic chart. First, assume that \( \tilde{S} \) is open, and that the triangulation is shellable, in the sense that there is an enumeration of the triangles \( \Delta_i \) so that \( D_n := \cup_{i \leq n} \Delta_i \) is topologically a disk, and \( \Delta_{n+1} \ldots \)
intersects this disk along a connected arc \( \gamma_n \) in the boundary, which is equal to one or two sides of the triangle.

Suppose by induction we have found a uniformizing map \( f_n : D_n \to D \) from \( D_n \) to the unit disk \( D \), conformal on the interior and extending continuously to the boundary. Let \( U_{n+1} \) be a chart containing \( \Delta_{n+1} \) such that \( g : U_{n+1} \to \mathbb{C} \) is a holomorphic embedding. We show how to obtain \( f_{n+1} : D_{n+1} \to D \). Perturb \( \gamma_n \) and the triangulation in the interior of \( D_{n+1} \) so that \( \gamma_n \) is a real analytic arc (this is only important if \( \Delta_{n+1} \) intersects \( D_n \) in two sides meeting in a corner). Let \( \sigma = f_n(\gamma_n) \), a round arc in the boundary of \( D \). By our hypothesis on \( \gamma_n \), the map \( f_n \) is real analytic on \( \gamma_n \), so we can apply the Schwarz reflection principle to extend \( f_n \) near \( \gamma_n \).

Explicitly, let \( B_1 \) be a bigon with round edges \( \sigma \cup \sigma_1 \) contained in \( D \) meeting at an angle \( 2^{-k}\pi \) for some big enough \( k \) so that \( f_{n+1}^{-1}(B_1) \subset U_{n+1} \), and let \( T_1 = \Delta_{n+1} \cup f_{n+1}^{-1}(B_1) \). Then \( g \) uniformizes \( T_1 \) as a topological disk in \( \mathbb{C} \), so by the Riemann mapping theorem we can find a uniformizing map \( g_1 : T_1 \to D \). Let \( B_2 \) be the bigon with round edges \( \sigma \) and \( \sigma_2 \) obtained by reflecting \( B_1 \) along \( \sigma_1 \), and let \( D_2 \) be obtained from \( D \) by reflecting \( g_1 f_{n+1}^{-1}(B_1) \) along the round arc \( g_1 f_{n+1}^{-1}(\sigma_1) \). Let \( T_2 = \Delta_{n+1} \cup f_{n+1}^{-1}(B_2) \), so that \( g_1 : T_2 \to D_2 \). We can further uniformize \( D_2 \) as \( D \), and denote the composition as \( g_2 : T_2 \to D \). Then obtain \( B_3 \) by reflecting \( B_2 \) along \( \sigma_2 \), obtain \( T_3 = \Delta_{n+1} \cup f_{n+1}^{-1}(B_3) \), obtain \( g_3 : T_3 \to D \), and so on. After \( k \) reflections, \( B_k \) is \( D \) so that \( T_k = \Delta_{n+1} \cup D_n = D_{n+1} \) and we have obtained a uniformizing map \( f_{n+1} : D_{n+1} \to D \) by setting \( f_{n+1} = g_k \).

We can normalize the \( f_n \) to take some fixed basepoint \( p \in \Delta_1 \subset \tilde{S} \) to 0, and satisfy \( f_n'(p) = 1 \) where the derivative is taken with respect to some (fixed) local holomorphic coordinate at \( p \). Then by the Schwarz Lemma, the \( f_n \) are locally bounded, and by Montel’s theorem they converge uniformly on compact subsets to some injective holomorphic map \( \tilde{S} \to \mathbb{C} \) whose image is either a round disk, or all of \( \mathbb{C} \). This completes the proof if \( \tilde{S} \) is open.

If \( \tilde{S} \) is closed, we can puncture it at a point and uniformize the complement. Because the modulus of a punctured disk is infinite, the uniformization has image all of \( \mathbb{C} \). Now extend over the puncture. \( \square \)

If \( S \) is a closed Riemann surface which is not a sphere, the deck group \( \pi_1(S) \) acts on \( \tilde{S} \) by conformal symmetries. Thus after uniformizing \( \tilde{S} \) as one of \( \mathbb{C} \) or \( D \), we obtain a discrete faithful cocompact representation of \( \pi_1(S) \) into \( \mathbb{C} \) (acting on \( \mathbb{C} \) by translations) or into \( \text{PSU}(1,1) \) (acting on \( D \) by hyperbolic isometries). In particular, \( \tilde{S} \) admits a unique Euclidean similarity structure (if \( S \) is a torus) or hyperbolic metric (if the genus of \( S \) is at least 2) in its conformal class.

4.3. Quasiconformal maps. A domain \( \Omega \) in \( \mathbb{C} \) is a real 2-dimensional manifold. If \( z := x + iy \) is a local holomorphic coordinate, then \( x, y \) are local real coordinates, and \( dz := dx + idy \) and \( d\bar{z} := dx - idy \) are a basis for the space of complex valued smooth 1-forms on \( \Omega \). Dual to these 1 forms are the complex valued smooth vector fields

\[
\frac{\partial}{\partial z} := \frac{1}{2} \left( \frac{\partial}{\partial x} - i \frac{\partial}{\partial y} \right), \quad \frac{\partial}{\partial \bar{z}} := \frac{1}{2} \left( \frac{\partial}{\partial x} + i \frac{\partial}{\partial y} \right)
\]

where we abbreviate \( \partial f/\partial z \) by \( f_z \) and \( \partial f/\partial \bar{z} \) by \( f_{\bar{z}} \) for a smooth complex-valued function \( f \). The Cauchy-Riemann equations say that \( f \) is holomorphic in \( z \) if and only if \( f_{\bar{z}} = 0 \).
For a smooth complex-valued function on Ω, the Jacobian \( J_f \) of \( f \) satisfies \( J_f = |f_x|^2 - |f_z|^2 \), so where \( f \) is locally an orientation-preserving diffeomorphism we have \( |f_x|^2 - |f_z|^2 > 0 \) or equivalently \( \mu := \frac{f_x}{f_z} \) has \( |\mu| < 1 \). If \( f \) is a smooth map between Riemann surfaces, we may choose holomorphic coordinates \( z \) on the domain and on the range, and then it makes sense to define the 1-forms \( f_x dz \) and \( f_z d\bar{z} \) and the Beltrami differential

\[
\mu_f(z) \frac{dz}{d\bar{z}} := \frac{f_x dz}{f_z d\bar{z}}
\]

We often suppress the \( f \) in the subscript of \( \mu \) 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 holomorphic 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

\[
K(z) = \frac{1 + |\mu(z)|}{1 - |\mu(z)|}
\]

A function \( f \) as above is said to be quasiconformal in a domain \( \Omega \) in a Riemann surface if \( K_f := \sup_{z \in \Omega} K(z) \) is finite. We call \( K_f \) the (maximal) dilatation of \( f \) in the domain \( \Omega \). Conformal maps are distinguished among all quasiconformal maps exactly by the condition \( K_f = 1 \), or equivalently \( \mu = 0 \).

4.3.1. 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 \( \epsilon \) there is a positive \( \delta \) 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 \( \nu \) 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 \( \Omega \) are absolutely continuous.

**Definition 4.6** (Analytic definition of quasiconformal maps). An orientation-preserving homeomorphism \( f : \Omega \to \mathbb{C} \) is quasiconformal if it satisfies

1. \( f \) is ACL on \( \Omega \); and
2. there exists a \( k \) with \( 0 \leq k < 1 \) such that \( |f_x| \leq k|f_z| \) a.e. on \( \Omega \).

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 \( f \).

4.3.2. Geometric definition. A quadrilateral is the closure of a domain \( Q \) bounded by a Jordan curve \( \partial Q \) together with 4 distinct points \( q_1 \cdots q_4 \) in cyclic order in \( \partial Q \). For each such quadrilateral there is a unique positive real number \( m \) (called the modulus of \( (Q, q_4) \)) for which there is a uniformizing map from \( Q \) to the rectangle \([0, 1] \times [0, m]\) taking the four points to the four vertices in cyclic order, with \( q_1 \) going to \((0, 0)\). Existence follows from the
Riemann mapping theorem (from $Q$ to the upper half-plane) followed by a suitable Schwarz-Christoffel map; uniqueness follows by the Schwarz reflection principle. Evidently cyclically permuting the vertices of a quadrilateral either preserves the modulus or interchanges it with $1/m$.

**Definition 4.7** (Geometric definition of quasiconformal maps). An orientation-preserving homeomorphism $f : \Omega \to \mathbb{C}$ is quasiconformal if there is some constant $K \geq 1$ such that $m(f(Q,q_*)) \leq Km((Q,q_*))$ for all quadrilaterals $(Q,q_*)$ in $\Omega$.

**Proposition 4.8** (Equivalence of definitions). The analytic and geometric definitions of quasiconformal maps are equivalent.

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.

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. Explicitly,

**Proposition 4.9** (Uniform convergence). Every sequence of $K$-quasiconformal maps of $\mathbb{C}$ onto itself fixing 0 and 1 contains a subsequence which converges uniformly in the spherical metric. Moreover, the limit function is also $K$-quasiconformal.

For proofs of Proposition 4.8 and Proposition 4.9 see Ahlfors [2].

4.3.3. *Geometry of 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 4.10.** A Jordan curve $\gamma$ in $\mathbb{C}$ has bounded turning if there is a constant $C$ such that if $z_1, z_2$ are chosen on $\gamma$, and $z_3$ is on the component of $\gamma - \cup z_i$ of least diameter, then

\[ |z_1 - z_3| + |z_2 - z_3| \leq C|z_1 - z_2| \tag{4.11} \]

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.

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

**Proposition 4.11.** 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 $\gamma$ is a geodesic, let $\delta$ be the geodesic obtained by straightening $F(\gamma)$. Now, suppose $\pi$ is a totally geodesic plane perpendicular to $\gamma$ at some point. We claim that there is a constant $C$ depending only on $K$ so that the projection of $F(\pi)$ to $\delta$ has diameter at most $C$. This implies in particular that the circle at infinity of $\pi$ 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 $\gamma'$ in $\pi$ intersecting $\gamma$, and let $\delta'$ be the geodesic obtained by straightening $F(\gamma')$. The image $F(\pi)$ is a bounded distance from the union of such $\delta'$, so it suffices to show that the projection of $\delta'$ to $\delta$ has bounded diameter. But this is equivalent to the condition that there is a bound on the length of segments in $\delta$ and $\delta'$ 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 $\gamma, \gamma'$.

Conversely, let $f$ be a quasiconformal homeomorphism of $S^2_\infty$. Douady and Earle define the following conformal barycenter extension of $f$, as follows. For each point $p \in \mathbb{H}^3$, let $\nu_p$ denote the visual measure on $S^2_\infty$ as seen from $p$. Then $\nu_p$ pushes forward to the probability measure $f_*\nu_p$ on $S^2_\infty$. 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 $S^2_\infty$ by the exponential map. Then define

\begin{equation}
V_p(q) = \int_{U_q\mathbb{H}^3} z \, d(f_*\nu_p)(z)
\end{equation}

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 [4] 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 $K$).

\begin{remark}
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 $\mu$, and think of $f$ as the time 1 map of a 1-parameter flow $f_t$ obtained by the Measurable Riemann Mapping Theorem (Theorem 4.13) 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 [9] B.4 for details.
\end{remark}

4.4. 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 [10], is as follows:

**Theorem 4.13 (Existence of quasiconformal homeomorphism).** For every measurable Beltrami differential \( \mu := \mu(z) dz/dz \) on \( \widehat{\mathbb{C}} \) with \( \text{ess sup}_z |\mu(z)| < 1 \), there is a quasiconformal homeomorphism \( f : \widehat{\mathbb{C}} \to \widehat{\mathbb{C}} \) with dilatation equal to \( \mu \) a.e. Moreover, \( f \) is unique if we further impose that \( f \) fixes 0, 1 and \( \infty \).

We sometimes denote the \( f \) associated to \( \mu \) promised by the theorem by \( f^\mu \). Ahlfors-Bers extended Theorem 4.13 to show that \( f \) depends holomorphically on \( \mu \), 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 \( \widehat{\mathbb{C}} \) associated to \( \mu \), 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 sufficiently regular, finding isothermal coordinates is not so hard; the real analytic content of the theorem is the existence of isothermal coordinates for such a metric when \( \mu \) is merely measurable.

The existence of isothermal coordinates for real analytic \( \mu \) is due to Gauss.

### 4.5. Teichmüller space.

**Definition 4.14 (Teichmüller space).** If \( S \) is a topological surface, the *Teichmüller space* of \( S \), denoted \( \mathcal{T}(S) \), is the space of equivalence classes of pairs \((R, f)\) where \( R \) is a Riemann surface and \( f : S \to R \) is a homotopy equivalence, and where \((R_1, f_1) \sim (R_2, f_2)\) if there is a holomorphic isomorphism \( g : R_1 \to R_2 \) so that \( gf_1 \) is homotopic to \( f_2 \).

The map \( f : S \to R \) is called a *marking*, and the data of the pair \((R, f)\) is called a *marked Riemann surface*. We have described Teichmüller space as a set; to describe it as a topological space it is easiest just to define a metric on it.

**Definition 4.15 (Teichmüller metric).** The *Teichmüller distance* between two marked Riemann surfaces \((R_1, f_1)\) and \((R_2, f_2)\) is defined to be \( \inf_g \log K(g) \) where \( K(g) \) is the 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.

Every conformal class of metric on a surface admits isothermal coordinates; i.e. it defines (uniquely) the structure of a compatible Riemann surface. If the genus of \( R \) is at least 2, the universal cover is biholomorphic to the unit disk \( D \) by the Uniformization Theorem.
Holomorphic automorphisms of $D$ are precisely the same as the orientation-preserving isometries in the Poincaré disk model. It follows that each such surface $R$ admits a unique hyperbolic metric in its conformal class. We may therefore equally well define $\mathcal{J}(S)$ to be the space of marked hyperbolic surfaces $(R, f)$ homeomorphic to $S$.

4.6. Kleinian groups.

**Definition 4.16** (Kleinian group). A Kleinian group $\Gamma$ is a discrete, finitely generated subgroup of $\text{PSL}(2, \mathbb{C})$.

Associated to a Kleinian group $\Gamma$ there is a natural decomposition of $S^2_\infty$ into two subsets canonically associated to $\Gamma$; the limit set $\Lambda(\Gamma)$, and the domain of discontinuity $\Omega(\Gamma)$. In the Poincaré ball model, the union $\mathbb{H}^3 \cup S^2_\infty$ is the closed unit ball. For any $x \in \mathbb{H}^3$ we can form the orbit $\Gamma x$ and take the closure $\overline{\Gamma x}$ in the closed unit ball. The limit set is then the difference $\Lambda(\Gamma) = \overline{\Gamma x} - \Gamma x = \overline{\Gamma x} \cap S^2_\infty$. Note that this set does not depend on the choice of point $x$.

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 4.17.** If $\Gamma$ is not elementary, $\Lambda$ is the unique minimal closed non-empty $\Gamma$-invariant subset of $S^2_\infty$.

**Proof.** Associated to any closed invariant subset $K$ we can form the convex hull $C(K)$. Since $\Gamma$ 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 $\Gamma x$ and (since it is closed) $\overline{\Gamma x}$ and therefore $\Lambda$. \hfill $\Box$

**Corollary 4.18.** If $\Gamma'$ is a normal subgroup of $\Gamma$ (both nonelementary), then $\Lambda(\Gamma') = \Lambda(\Gamma)$.

**Proof.** Since $\Gamma$ conjugates $\Gamma'$ to itself, it takes $\Lambda(\Gamma')$ to itself. Since $\Gamma'$ is nonelementary, $\Lambda(\Gamma')$ is nonempty. \hfill $\Box$

**Example 4.19.** Let $\Gamma$ be the group generated by a surface subgroup stabilizing a totally geodesic plane $\pi$ in $\mathbb{H}^3$, together with reflection in $\pi$. The reflection generates a nontrivial normal subgroup $\Gamma'$ which is nevertheless elementary (this contradicts [13], 8.1.3).

**Example 4.20.** 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 $S \rightarrow M$ denotes inclusion of the fiber, there is a short exact sequence of fundamental groups

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

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 and two ends (at least if $M$ is compact). These are examples of geometrically infinite ends.

A finitely generated Kleinian group has only finitely many conjugacy classes of torsion elements, and has a torsion-free subgroup of finite index. For simplicity in the sequel
we consider only torsion-free Kleinian groups. For such a group, the quotient $\mathbb{H}^3/\Gamma$ is a complete hyperbolic 3-manifold (typically of infinite volume).

The action of $\Gamma$ on $\mathbb{H}^3 \cup \Omega$ is properly discontinuous; one way to see this is to form the convex hull $C(\Gamma)$ of the limit set and observe that nearest point projection gives a canonical deformation retraction from $\mathbb{H}^3 \cup S^2_\infty$ to $C(\Gamma)$ taking $\Omega$ to $\partial C \cap \mathbb{H}^3$. This projection commutes with the action of $\Gamma$, which is properly discontinuous on $C \cap \mathbb{H}^3$, and the claim follows. Consequently we can form the Kleinian manifold quotient $N(\Gamma) := \mathbb{H}^3 \cup \Omega/\Gamma$. This is a 3-manifold with some boundary components $\partial N$ obtained as a quotient $\Omega/\Gamma$ (of course if $\Omega$ is empty then $N = M$).

4.7. Geometrically finite manifolds. We are using the notation $C(\Gamma)$ or $C$ for the (closed) convex hull of the limit set of a Kleinian group $\Gamma$ in $S^2_\infty \cup \mathbb{H}^3$. We use the notation $\hat{C}(\Gamma)$ (or just $\hat{C}$) for $C(\Gamma) \cap \mathbb{H}^3$, and $M_C(\Gamma)$ (or just $M_C$) for the quotient $\hat{C}/\Gamma$. We refer to $M_C$ as the convex hull of $M$.

**Definition 4.21.** A Kleinian group is geometrically finite if the $\epsilon$-neighborhood of the convex hull $N_\epsilon(M_C) \subset M$ has finite volume.

**Theorem 4.22** (Ahlfors). If $\Gamma$ is geometrically finite then $\Lambda$ has either zero measure or full measure. If $\Lambda$ has full measure, the action of $\Gamma$ on $S^2_\infty$ is ergodic.

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

**Proof.** It is equivalent to show that any bounded measurable function $f$ supported on $\Lambda$ and invariant by $\Gamma$ is constant a.e. with respect to Lebesgue measure. Let $f$ be such a function, and let $h_f$ be the harmonic extension to $\mathbb{H}^3$; i.e. the function which at every point $x$ in $\mathbb{H}^3$ is equal to the average of $f$ with respect to the visual measure on $S^2_\infty$ as seen from $x$. Without loss of generality we may assume $f$ takes only the values $0$ and $1$. Now, if $x$ is outside $C$ then $h_f(x) < 1/2$, since there is a hemisphere containing $x$ and missing $\Lambda$. Since $f$ is $\Gamma$-invariant, so is $h_f$, so it descends to a function (which by abuse of notation we also call $h_f$) on $M$. Now, the subset where $h_f \geq 1/2$ is contained in $M_C$ which has finite volume. Since $h_f$ is harmonic, its gradient flow is volume preserving; but this flow takes the subset where $h_f \geq 1/2$ inside itself, which gives a contradiction unless $h_f < 1/2$ 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 $1$. \qed

**Example 4.23** (Convex cocompact). The simplest example of a geometrically finite Kleinian group $\Gamma$ is one for which the convex hull $M_C$ is compact; such $\Gamma$ are said to be convex cocompact. In this case, nearest point retraction gives an isomorphism between $\partial N$ and $\partial M_C$, so that $N$ is also compact, and $M$ is homeomorphic to the interior of $N$. A special case, of course, is when $M$ is closed.

**Example 4.24.** Suppose $\Gamma$ is discrete and infinitely generated, and stabilizes a totally geodesic $\mathbb{H}^3$ in $\mathbb{H}^3$ with limit set a round circle, and such that the quotient $\mathbb{H}^3/\Gamma$ has infinite area. Then the convex hull of $\mathbb{H}^3/\Gamma$ is 2-dimensional, and has zero volume, but a neighborhood $N_\epsilon(M_C)$ has infinite volume.
Suppose \( \Gamma \) is convex cocompact. In this case, the neighborhood \( N_\epsilon(M_\mathcal{C}) \) 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 \( N_\epsilon(M_\mathcal{C}) \) will still be uniformly strictly convex, and will therefore contain the convex hull for the deformed structure (which must, \textit{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.

\textit{Example} 4.25 (Thurston’s mickey mouse). Let \( R \) be a hyperbolic surface of genus 2, and let \( \Gamma \) be a Kleinian group stabilizing \( \pi \), a totally geodesic \( \mathbb{H}^2 \) in \( \mathbb{H}^3 \), with quotient \( R \). Let \( \gamma \) be a simple closed geodesic on \( R \); for simplicity, suppose \( \gamma \) separates \( R \) into \( R_1 \) and \( R_2 \), surfaces with boundary. The fundamental group of \( R \) is an amalgam

\begin{equation}
\pi_1(R) = \pi_1(R_1) \ast_\langle \gamma \rangle \pi_1(R_2)
\end{equation}

and we obtain \( \Gamma \) as the image of a discrete faithful representation \( \rho : \pi_1(R) \to \text{PSL}(2, \mathbb{C}) \). Now, let \( \ell \) be the axis of \( \rho(\gamma) \) in \( \mathbb{H}^3 \), and let \( \alpha(\theta) \) be an elliptic element which rotates through angle \( \theta \) about the axis \( \ell \). We can build a new representation \( \rho_\theta : \pi_1(R) \to \text{PSL}(2, \mathbb{C}) \) which agrees with \( \rho \) on \( \pi_1(R_1) \), and which conjugates \( \rho(\pi_1(R_2)) \) by \( \alpha(\theta) \). Since \( \alpha(\theta) \) commutes with \( \rho(\gamma) \), these representations agree on \( \gamma \), and piece together to give a representation of \( \pi_1(R) \).

It is easier to think about the geometry of the universal cover. The group \( \rho(\pi_1(R)) \) \( \pi \) and all the translates of \( \ell \) cut \( \pi \) up into subsurfaces. We take \( \pi \) and bend it simultaneously along each translate of \( \ell \) through the same angle \( \alpha \). The result is a plane \( \pi_\alpha \) which has been “pleated” along these geodesics. We claim that for fixed \( \gamma \) there is an angle \( \alpha_0 > 0 \) so that for \( \alpha < \alpha_0 \) the plane \( \pi_\alpha \) is quasi-isometrically embedded in \( \mathbb{H}^3 \), and limits to a quasicircle. This is for the following reason. If \( r \) is the length of the shortest (homotopically nontrivial) geodesic arc from \( \gamma \) to itself, then any geodesic \( \delta \) in \( \pi \) gets bent to a piecewise geodesic \( \delta' \) in \( \pi' \) which is made up of segments of length at least \( r \) bent along angles of at most \( \alpha \).

But providing \( \sin(\alpha/2) < r/2 \) this piecewise geodesic makes linear progress on large scales, and is therefore a quasigeodesic. Thus the normals from \( \pi_\alpha \) ultimately diverge, and give an isomorphism between \( \pi_\alpha \) and the two components of the domain of discontinuity.

\textit{Example} 4.26. Let \( C_1, \ldots, C_n \) be a union of round circles with disjoint interiors, tangent to each other in a ring around another round circle \( C \) to which they are all orthogonal. Inversion in each \( C_i \) preserves \( C \), and these inversions generated a Kleinian group \( \Gamma \) with \( C \) as the limit set. As we slide the \( C_i \) around, still keeping them tangent with disjoint interiors, but no longer simultaneously orthogonal to some circle, the limit set deforms to a quasicircle. When two non-adjacent (in the cycle) circles touch, the quasicircle pinches off into a cactus. See Figure 7.

\textit{Example} 4.27 (Bers simultaneous uniformization). Let \( S \) be a closed oriented surface of genus at least 2, and let \( (R_1, R_2) \in \mathcal{T}(S) \times \mathcal{T}(\hat{S}) \) be a pair of conformal structures on \( S \) and \( \hat{S} \). Then there is a discrete faithful representation \( \rho : \pi_1(S) \to \text{PSL}(2, \mathbb{C}) \) whose limit set is a quasicircle dividing \( \hat{C} \) into two topological disks, and the quotient of these disks by the action of \( \pi_1(S) \) gives rise to the (marked) Riemann surfaces \( R_1, R_2 \).

To see this, first uniformize the universal cover of \( R_1 \) as the upper half-plane, obtaining a Fuchsian group \( \Gamma \) with limit set \( \hat{\mathbb{R}} \) dividing \( \hat{C} \) into two topological disks with quotients
Figure 7. Quasicircles obtained by inversions in a ring of tangent circles. Degeneration to a cactus when two nonadjacent circles become tangent.

$R_1$ and $\overline{R}_1$. Choose a diffeomorphism $\overline{R}_1 \to R_2$ in the correct homotopy class, and lift the Beltrami differential to a differential $\mu$ on the lower half-plane, and extend by zero over the upper half-plane. Then apply the measurable Riemann mapping theorem.

There are some equivalent characterizations of geometric finiteness which give the concept more substance:

**Proposition 4.28.** The following are equivalent for $\Gamma$ a discrete, torsion-free subgroups of $\text{PSL}(2, \mathbb{C})$:

1. $\Gamma$ is geometrically finite;
2. the $\epsilon$-thick part of the convex hull is compact for any $\epsilon$; and
3. $\Gamma$ admits a finite-sided fundamental domain.

**4.8. Ahlfors finiteness theorem.** Recall if $\Gamma$ is a discrete subgroup of $\text{PSL}(2, \mathbb{C})$, we can decompose $S^2_\infty$ into the limit set $\Lambda$ and the domain of discontinuity $\Omega$. The quotient $\Omega/\Gamma$ is a (possibly disconnected) (marked) Riemann surface which is an invariant of the conjugacy class of $\Gamma$ (when $\Gamma$ has torsion, $\Omega/\Gamma$ might have the natural structure of an orbifold; but we elide this possibility from the discussion for simplicity).

When $\Gamma$ is finitely generated, Ahlfors showed that $\Omega/\Gamma$ cannot be too complicated. A Riemann surface is said to be *analytically finite* if it is isomorphic to a closed Riemann surface minus finitely many points.

**Theorem 4.29 (Ahlfors finiteness).** Let $\Gamma$ be a finitely generated torsion-free Kleinian group. Then $\Omega/\Gamma$ is a finite union of analytically finite Riemann surfaces.

The proof we give here follows an argument of Sullivan and Bers, but is in the same spirit as Ahlfors’ original argument. There was a lacuna in Ahlfors’ original proof concerning thrice-punctured spheres, which can be finessed by an appeal to Selberg’s lemma (although there are several alternate methods to fill the gap).

If $\Gamma$ is a discrete subgroup of $\text{PSL}(2, \mathbb{C})$, a *Beltrami differential* for $\Gamma$ is an essentially bounded measurable $\mathbb{C}$-valued function $\mathcal{C}$ on $\hat{\mathbb{C}}$ which transforms as

$$\mu(\gamma(z)) \overline{\gamma'(z)} = \mu(z)$$

for all $\gamma \in \Gamma$ and all $z$. Equivalently, the $(-1, 1)$-differential $\mu(z)dz/dz$ is invariant by $\Gamma$. Define the Banach space $B(\Gamma)$ of Beltrami differentials for $\Gamma$, and denote by $B_1(\Gamma)$ the open unit ball; i.e. the set of Beltrami differentials with $\|\mu\|_\infty < 1$. The measurable Riemann
mapping theorem associates to each \( \mu \in B_1(\Gamma) \) a quasiconformal automorphism \( f \) of \( \hat{\mathbb{C}} \) with \( \mu = \mu_f \), unique up to composition with a Möbius transformation.

A holomorphic quadratic differential for \( \Gamma \) on a domain \( \Omega \) is a holomorphic \((2,0)\)-differential on \( \Omega \) invariant under \( \Gamma \); i.e. it is given in a local coordinate \( z \) by a holomorphic function \( \varphi(z) \) which transforms as

\[
\varphi(\gamma(z))\gamma'(z)^2 = \varphi(z)
\]

for all \( \gamma \in \Gamma \) and all \( z \). If \( \varphi \) is a holomorphic quadratic differential on a conformally hyperbolic domain \( \Omega \), and if \( \lambda_0|dz| \) is the hyperbolic metric on \( \Omega \), then we define a norm on \( \varphi \) by

\[
\|\varphi\| = \sup_{z \in \Omega} \lambda_0^{-2}(z)|\varphi(z)|
\]

Note that this expression is independent of the choice of holomorphic coordinate \( z \). If \( \Gamma \) is nonelementary, the domain of discontinuity \( \Omega \) is conformally hyperbolic, and we can define the Banach space of holomorphic quadratic differentials for \( \Gamma \) with finite norm, and denote it by \( A^\infty(\Omega, \Gamma) \). Because of the automorphic properties of such differentials, they descend to well-defined holomorphic quadratic differentials on \( \Omega/\Gamma \). In other words, we can identify \( A^\infty(\Omega, \Gamma) \) with \( A^\infty(\Omega/\Gamma) \).

For a Riemann surface \( R \), realized as \( \mathbb{H}^2/\Gamma \), there is a projection from \( B_1(\mathbb{H}^2, \Gamma) \) to \( A^\infty(\overline{R}) \), called the Bers projection. For \( \mu \in B_1(\mathbb{H}^2, \Gamma) \), identify \( \mathbb{H}^2 \) with the upper half-plane, to obtain a Beltrami differential \( \hat{\mu} \) there invariant by \( \Gamma \), and extend it by zero in the lower half-plane. Let \( f \) be the quasiconformal automorphism of \( \hat{\mathbb{C}} \) with differential \( \hat{\mu} \). The restriction of \( f \) to the lower half-plane is conformal, and its Schwarzian \( S(f) \) is a holomorphic quadratic differential on the lower half-plane invariant under \( \Gamma \). Since the quotient of the lower half plane by \( \Gamma \) is isomorphic to \( \overline{R} \), we have \( S(f) \in A^\infty(\overline{R}) \). Define \( \Phi(\mu) = S(f) \); this is the Bers projection.

**Proposition 4.30.** The Bers projection \( \Phi \) is a holomorphic open map from \( B_1(R) \) onto a bounded set in \( A^\infty(\overline{R}) \). Moreover, \( \Phi(\mu) = \Phi(\nu) \) if and only if the quasiconformal automorphisms of \( R \) determined by \( \mu \) and \( \nu \) produce isomorphic (marked) Riemann surfaces.

There is an explicit holomorphic section \( \sigma : A^\infty(\overline{R}) \to B_1(R) \) for \( \Phi \), given by

\[
\sigma(\varphi)(z) = -2\lambda_H^{-2}(z)\varphi(\overline{z})
\]

This takes a quadratic holomorphic differential of norm less than \( 1/2 \) to a Beltrami differential of norm less than \( 1 \). Beltrami differentials in \( B_1(R) \) of the form \( \lambda^2(z)\overline{\varphi}(z) \) are called harmonic Beltrami differentials for \( \Gamma \) on \( \Omega \). They minimize the dilatation amongst all differentials associated to quasiconformal maps whose image is a fixed Riemann surface.

To prove the Ahlfors finiteness theorem, we must now just count dimensions. For simplicity, suppose all components of \( \Omega \) are simply-connected (the general case is not much harder to handle, and is best treated by a mix of analytic and topological methods). A component of \( \Omega/\Gamma \) which is not analytically finite admits an infinite dimensional space of holomorphic quadratic differentials. An analytically finite component of genus \( g \) with \( n \) punctures satisfies

\[
\dim(A^\infty) = 3g - 3 + n
\]
as can be seen from the Riemann-Roch formula. Associated to these families of differentials are families of harmonic Beltrami differentials for \( \Gamma \) supported in \( \Omega \), which by the measurable Riemann mapping theorem define families of nontrivial deformations of the representation of the abstract group \( \Gamma \) into \( \text{PSL}(2, \mathbb{C}) \) (up to conjugacy). Since \( \Gamma \) is finitely generated, the space of representations is finite (complex) dimensional (at most 3 times the number of generators). Thus \( \Omega/\Gamma \) consists of finitely many analytically finite Riemann surfaces — plus possibly infinitely many thrice punctured spheres (for which \( A^\infty \) is trivial). The last step is to rule out this possibility.

4.8.1. Busting thrice-punctured spheres. If \( \Gamma' \) is a finite index subgroup of \( \Gamma \), then they have the same domain of discontinuity, and \( \Omega/\Gamma' \) is a finite cover of \( \Omega/\Gamma \). Thus to prove Ahlfors finiteness for \( \Gamma \) it suffices to prove it for \( \Gamma' \). We now explain how to use an algebraic trick to find a subgroup \( \Gamma' \) of finite index for which one can guarantee than \( \Omega/\Gamma' \) contains no thrice punctured spheres at all. This will complete the proof.

We first start with a lemma, which shows that thrice-punctured spheres in hyperbolic 3-manifolds are rigid — i.e. they have no moduli, and are always totally geodesic.

**Lemma 4.31 (Thrice-punctured sphere rigid).** Let \( \Gamma \) be nonelementary Kleinian generated by two parabolic elements \( f \) and \( g \) such that \( fg \) is parabolic. Then \( \Gamma \) is conjugate to the group

\[
\Gamma \sim \left\langle \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix} \right\rangle
\]

*Proof.* The elements \( f \) and \( g \) have distinct fixed points, which we can take to be \( \infty \) and \( 0 \) in the upper half-space model. After conjugacy we can assume \( f(z) = z + 2 \) and \( g(z) = \frac{z}{cz + 1} \) for some constant \( c \). Since \( fg \) is parabolic, we compute \( c = -2 \). □

It follows that every thrice-punctured sphere in \( \Omega/\Gamma \) contains a conjugacy class with specific trace \( \neq 2 \); e.g. trace 6. On the other hand, we have the following algebraic fact, a variant on what is sometimes known as Selberg's Lemma:

**Lemma 4.32.** Let \( G \) be a finitely generated subgroup of \( \text{SL}(2, \mathbb{C}) \) and let \( t \neq 2 \). Then there exists a finite index normal subgroup \( N \) of \( G \) such that no element of \( N \) has trace \( t \).

*Proof.* Let \( R \) be the ring generated by the matrix entries of the generators of \( G \) 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 zero, 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 \( \phi : \text{SL}(2, R) \to \text{SL}(2, K) \) obtained by reducing entries mod \( \mathfrak{m} \), and the kernel intersects \( G \) 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 \( \text{tr}(A) = t - 2 \). But \( A \) has entries in \( \mathfrak{m} \), so \( \text{tr}(A) \) is in \( \mathfrak{m} \), contrary to the hypothesis that \( t - 2 \) is not in \( \mathfrak{m} \). Thus \( N \) is the desired subgroup. □

Thus for any finitely generated Kleinian group \( \Gamma \), we can always find a finite index subgroup \( \Gamma' \) for which \( \Omega/\Gamma' \) contains no thrice-punctured spheres, and is therefore (by the argument above) a finite union of analytically finite surfaces. And so therefore is \( \Omega/\Gamma \), proving Theorem 4.29.
4.9. **No invariant line fields.** Proposition 4.11 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 \( \Omega \) are parameterized by the Teichmüller spaces of \( \Omega/\Gamma \), which by Theorem 4.29 is finite dimensional. But this leaves open the possibility that there might be Beltrami differentials \( \mu \) for \( \Gamma \) supported in the limit set. For *geometrically finite* \( \Gamma \), this is impossible, since Theorem 4.22 says in this case either the limit set has zero measure (in which case it can’t support a Beltrami differential), or it is full.

Sullivan [12] proved the following theorem:

**Theorem 4.33** (Sullivan; no invariant line field). *Let \( \Gamma \) be a finitely generated Kleinian group. Then any Beltrami differential for \( \Gamma \) vanishes a.e. on \( \Lambda \).*

5. **Hyperbolization for Haken manifolds**

5.1. Circle packing.

5.2. Andreev’s theorem.

5.3. Orbifold trick and reduction to the last step.

5.4. Skinning map.

5.5. Bounded image theorem.

5.6. Only windows break.

5.7. Double limit theorem.

5.8. Fibered case.

6. **Tameness**


6.2. Bonahon’s exiting sequences.

6.3. Shrinkwrapping.

6.4. Harmonic functions.

6.5. Ahlfors’ Conjecture.

7. **Ending laminations**

8. **Acknowledgments**

Danny Calegari was supported by NSF grant DMS 1405466.
NOTES ON KLEINIAN GROUPS

REFERENCES


University of Chicago, Chicago, Ill 60637 USA
E-mail address: dannyc@math.uchicago.edu