Math 6640 – Hyperbolic Geometry
Course Notes, Fall 2023

Lecturer: Benjamin Dozier

December 7, 2023

1 Motivations for hyperbolic geometry

Euclid's postulates:

1. Any two points can be joined by a unique line segment.
2. Any line segment can be extended indefinitely in both directions.
3. Given any line segment, a circle can be drawn with center at one endpoint, and with the other endpoint lying on the circle.
4. Any two right angles are congruent.
5. ("Parallel Postulate") If a line ℓ intersects two other lines m_1, m_2 and the sum of the angles made on one side of ℓ sum to less than 180 degrees, then m_1 and m_2 meet at some point.

Certain terms such as point, line, congruent are not defined by the postulates.

People spent literally thousands of years trying to prove the 5th postulate from the other 4 (as well as some "common notions" and implicitly used axioms about notions like "betweenness"). There was a good reason they failed: it is not possible, as the example of hyperbolic geometry would show. This was one of the great intellectual surprises in history.

Closed surfaces. Compact orientable surfaces (without boundary) are classified by *genus*, which takes values $0, 1, 2, \ldots$. On the sphere (genus 0), one can put a very nice geometric structure, the "round" sphere, coming from the standard embedding in $\mathbb{R}^3$ as the locus of solutions to $x^2 + y^2 + z^2 = 1$. The torus has an infinite (2 real dimensional) family of nice geometries; each comes from gluing pairs of parallel sides of some parallelogram. The geometry of the sphere is *positively curved*; the geometries on the torus have *zero curvature*. Are there similar "natural" geometries on surfaces of higher genus?

2 First models of hyperbolic geometry

2.1 Upper half-space model of hyperbolic plane

- Space: $\mathbb{H} := \{x + iy : y > 0\} \subset \mathbb{C}$.
- Riemannian metric: $ds^2 = \frac{1}{y^2}(dx^2 + dy^2)$.
- Geodesics: Half circles that meet the real axis perpendicularly, and vertical lines.

  One sees quickly that there exists a unique such object through any two distinct points (for two points not on the same line, take the perpendicular bisector of the segment formed by the two points. The intersection of this line with the real axis is the center of the desired circle).

  To see that these are in fact geodesics, start with the vertical lines, which are easier to deal with. Any path between two points on the same vertical line can be projected to the line without decreasing length. In fact, this argument shows that this is the only geodesic segment connecting two points on the same vertical line.

  The maps $\mathbb{H} \to \mathbb{H}$ given by $z \mapsto \lambda z$ for $\lambda > 0$, $z \mapsto z + c$ for $c \in \mathbb{R}$, and $z \mapsto -1/z$ are all seen to preserve the metric, and hence are *isometries*. For instance, the pullback of the arc-length element $|ds| = (1/\operatorname{Im}(z))|dz|$ along $z \mapsto -1/z$ is, using the identity $\operatorname{Im}(1/z) = -\operatorname{Im}(z)/|z|^2$:

  $$\frac{1}{\operatorname{Im}(1/z)}|d(1/z)| = \frac{-|z|^2|}{\operatorname{Im}(z)}\frac{-1}{|z|^2}|dz| = \frac{1}{\operatorname{Im}(z)}|dz|.$$

  The group generated by these acts transitively on the set of objects we claim are geodesics. Since vertical lines are geodesic, and isometries preserve geodesics, the rest are too.

- Isometry group. The isometries used above generate the group

  $$\left\{\frac{az+b}{cz+d} : a, b, c, d \in \mathbb{R}, ad - bc = 1\right\} \cong PSL_2(\mathbb{R}).$$

  This is exactly the group of *orientation preserving isometries* of $\mathbb{H}$. The full group of isometries is obtained by adding one additional generator: the map $z \mapsto -\bar{z}$. To see this is the full group of isometries, one can use that in a connected Riemannian manifold, an isometry is determined by what it does to a frame, i.e. a basis for the tangent space at a point.

Inversions. Given a circle C in the Euclidean plane, with center p and radius r, the *inversion* i_C about C is the map that takes a point $z \neq p$ to z' on the same line through 0 such that $|pz| \cdot |pz'| = r^2$. The point p is thought of as mapping to ∞.

Note that i_C is the identity on C. It fixes setwise any circle orthogonal to C (by power of a point). Inversion is *conformal*: it preserves angles (given two tangent vectors at a point not on C, construct circles A_1, A_2 through them and orthogonal to C. The image of the two tangent vectors will be at the other intersection of A_1, A_2, and the angle is the same here).

Inversion through the unit circle is given by the formula $z \mapsto \overline{1/z}$. This map (restricted to $\mathbb{H}$) is a composition of two of the isometries discussed above, and is hence itself an isometry for the hyperbolic metric.

Proposition 2.1. *Inversion i_C takes circles/lines to circles/lines.*

Proof (from Thurston book). We begin with the family $\mathcal{T}$ of circles/lines that are *tangent* to the circle C at some $z \in C$. These fill the plane, and they define a tangent line field $V_{\mathcal{T}}$ on the plane. We already know the family $\mathcal{O}$ of circles/lines *orthogonal* to C and passing through z is preserved; this family gives another line field $V_{\mathcal{O}}$. Note that $V_{\mathcal{T}}$ and $V_{\mathcal{O}}$ are orthogonal (at some other point z', the lines are tangent to some circle/line $C_1 \in \mathcal{T}$ and $C_2 \in \mathcal{O}$; these intersect at right angle at z, hence also at z').

Now consider the image under i_C of $V_{\mathcal{T}}$. Since i_C is conformal, $i_C(V_{\mathcal{T}})$ is orthogonal to $i_C(V_{\mathcal{O}}) = V_{\mathcal{O}}$. Since $V_{\mathcal{T}}$ is also orthogonal to $V_{\mathcal{O}}$, we must have that $i_C(V_{\mathcal{T}}) = V_{\mathcal{T}}$. And $i_C(V_{\mathcal{T}}) = V_{i_C(\mathcal{T})}$. The family of curves is determined by its tangent line field, so from $V_{\mathcal{T}} = V_{i_C(\mathcal{T})}$ we get that $i_C(\mathcal{T}) = \mathcal{T}$.

We can reduce to the tangent case by applying a scaling through the center of C (homothety), which interacts nicely with inversion. ∎

Such maps are called *Möbius transformations.*

Failure of parallel postulate. Take m_1 (resp. m_2) a semicircle centered at 0 and going through $2i$ (resp. i). Take ℓ the vertical line through 0. This situation does not violate the parallel postulate; m_1 and m_2 do not intersect, but the angles do not satisfy the required strict inequality. However, we deform m_2 slightly so that it goes through i but does make a right angle there, then we do get a violation of the parallel postulate.

2.2 Poincaré disc model

- Space: $\mathbb{D} := \{z : |z| < 1\} \subset \mathbb{C}$.
- Riemannian metric: $ds^2 = \frac{4}{(1-|z|^2)^2}(dx^2 + dy^2)$.
- Geodesics: circular arcs that meet the unit circle perpendicularly, and diameters of the unit circle.

  To find such an object connecting two distinct points $w, z \in \mathbb{D}$, we use inversion. Begin by constructing the point $i_C(w)$, where C is the unit circle. We claim that circle C' through $w, z, i_C(w)$ meets $\mathbb{D}$ in the desired arc (if w, z lie on the same diameter, these three points do not form a triangle; in this case that diameter is the required object). To see that C' is perpendicular to C, use power of a point.

- Isometries. Rotation about 0 through any angle is clearly an isometry. Others are harder to see.

2.3 Relation between half-space and disc models

We can find an inversion that takes the unit circle to a line (union $\{\infty\}$). Composing this with an affine transformation, we can map this line to the real axis. Composing all of this with a reflection if necessary gives a map $\mathbb{D} \to \mathbb{H}$ that preserves circles/lines. The map is also conformal, so it takes arcs/segments perpendicular to unit circle to arcs/segments perpendicular to real axis. That is, the map takes geodesics to geodesics.

One example of such a map $\phi : \mathbb{D} \to \mathbb{H}$, written in complex coordinates is,

$$\phi(z) := i \cdot \frac{1-z}{1+z}.$$

To see that the disc and upper-half space models are isometric other, we pull back the metric on $\mathbb{H}$ by the map ϕ above, using that $\mathrm{Im}(\phi(z)) = \mathrm{Re}\left(\frac{(1-z)(1+\bar{z})}{(1+z)(1+\bar{z})}\right) = \frac{1-|z|^2}{|1+z|^2}$:

$$\phi^*\left(\frac{1}{\mathrm{Im}(z)}|dz|\right) = \frac{1}{\mathrm{Im}(\phi(z))}|d\phi(z)| = \frac{|1+z|^2}{1-|z|^2} \cdot \left|\frac{2}{(1+z)^2}\right| |dz| = \frac{2}{1-|z|^2}|dz|.$$

Thus the map ϕ is an isometry. This means that intrinsic geometric properties are really the same for both models.

Nevertheless the models are useful for different things. Real scaling and translation are natural isometries of $\mathbb{H}$. There are corresponding isometries of $\mathbb{D}$, but they are a little harder to see. On the other hand, rotation about the origin is a natural isometry of $\mathbb{D}$; this is harder to see in $\mathbb{H}$.

2.4 Boundary at infinity

The boundary line of $\mathbb{H}$ is not part of the space itself, nor is the unit circle part of $\mathbb{D}$. Nevertheless, these objects have intrinsic geometric significance. The boundary $\partial\mathbb{H} := \mathbb{R} \cup \{\infty\}$ is augmented with the point ∞, while the boundary $\partial\mathbb{D} := S^1$ is already big enough. We get spaces $\overline{\mathbb{H}} := \mathbb{H} \cup \partial\mathbb{H}$, and $\overline{\mathbb{D}} := \mathbb{D} \cup \partial\mathbb{D}$; these come with natural topologies.

These are "boundaries at infinity". They can be defined intrinsically as the space of oriented geodesics, modulo the equivalence relation that two geodesics are equivalent if they become arbitrarily close in the forward direction. That is we think of the boundary as "endpoints" of the (infinitely long) geodesics.

2.5 Hyperbolic circles

The circle $S(x, r)$ is defined as all the points at distance r (measured in hyperbolic metric) from x.

The circle $S(0,r)$ in $\mathbb{D}$ is clearly just a Euclidean circle (of some possibly different radius), since the metric is radially symmetric. By applying an isometry, we can move any point to x. Isometries preserve Euclidean circles/lines, so the image must also be a Euclidean circle. But centers are not preserved.

2.6 Equidistant curves to geodesics

Instead of thinking of the locus of points equidistant to some point x, we can replace x by a geodesic γ and define

$$E(\gamma, r) := \{z \in \mathbb{H} : d(\gamma, z) = r\}.$$

To understand what this looks like, it is easiest to work in the upper-half space model and take γ to be a vertical line, say through 0. We claim then that $E(\gamma, r)$ is a Euclidean straight line L through 0. In fact, both γ and L are preserved by any isometry $z \mapsto \lambda z$, $\lambda \in \mathbb{R}$. Since this acts transitively on each, the curves must be equidistant.

By applying isometries (eg composition of inversions), we see that in general, in either disc or half-space model, an equidistant curve to a geodesic is a Euclidean arc (or line) meeting the boundary at two points.

We can also use this to see why it is plausible that the metric should have the form that it does given that we want the isometry group to contain the examples we've seen. The length ℓ of the circular arcs between γ and L meeting them perpendicularly should all have the same hyperbolic length, and this should only depend on the angle α at which γ meets L. In the limit as $\alpha \to 0$, it should be an approximately linear function of α, and we can take $d\ell/d\alpha = 1$. Now to find the length of a tangent vector v at some point, we can assume that it is a horizontal pointing vector in the upper-half space model, and then we approximate it by a circular arc between such a γ, L meeting at angle α. For v, small the Euclidean length of v is approximately $y \sin\alpha \approx y\alpha$. Since α is approximately the hyperbolic distance, we see where the factor of $1/y$ comes from in the expression for the hyperbolic metric.

2.7 Inversion in higher dimensions

The definition of inversion generalizes naturally to any dimension. Given a sphere $S = S^{n-1}$ in Euclidean space $\mathbb{R}^n$, with center p and radius r, the *inversion* i_S about S is the map that takes a point $z \neq p$ to z' on the same line through 0 such that $|pz| \cdot |pz'| = r^2$. The point p is thought of as mapping to ∞. (So we are thinking of $\mathbb{R}^n \cup \{\infty\}$ as S^n.)

Any sphere R orthogonal to S is preserved as a set by i_S. Using this, one can show (as in dimension 2) that inversions are conformal.

2.8 Stereographic projection

Given a sphere $S = S^{n-1}$ in $\mathbb{R}^n$ centered at p and an $(n-1)$ plane P tangent to S, we define a *stereographic projection* map $\phi : S - N \to P$, where N ("north pole") is antipodal to the point of tangency. For $z \in S - N$, we draw the line connecting z and N, and define $\phi(z)$ to be the intersection of this line with P.

We claim that ϕ is equal to the inversion $i_{\tilde{S}}$, where $\tilde{S}$ is the sphere centered at N with twice the radius of S. To see this, we use that line throughs the center N are fixed setwise by $i_{\tilde{S}}$, and a sphere through N and tangent to S' is taken to a plane tangent to S' at the same point (so S maps to P).

Since inversions are conformal, so is stereographic projection. And circles/lines are preserved (since higher inversions preserve spheres/planes).

We can also perform stereographic projection from $S - N$ to any plane P' parallel to P (other than the plane through N). By identifying P with P' via vertical translation, we can think of this map as stereographic projection to P composed with a dilation. Hence these more general stereographic projections are also conformal and preserve spheres/planes

Upper hemisphere model: project equatorial plane from south pole.

2.9 Upper hemisphere and Klein models

If we think of S a sphere, P a horizontal equatorial plane, and N the *south* pole, then there is an inverse stereographic projection taking the P to S. We think of the Poincaré disc D embedded as the equatorial disc, and then the map restricts to a map $f : D \to U$, where U is the open upper hemisphere of S. By transferring the metric on D via this map f, we get a hyperbolic model with space U; this is the *upper hemisphere* model.

Since we get from D to U by restricting inverse stereographic projection (which is conformal and preserves circles/line), the geodesics in U are arcs perpendicular to the equator. These are exactly the intersections of vertical planes with the upper hemisphere U.

We get the *Klein model* K (also known Beltrami-Klein model, or projective model) by taking the *orthogonal* projection from U to the equatorial disc. By the above description of geodesics in U, we see easily that the geodesics in the Klein model are just straight line segments connecting boundary points. So geodesics in K are simple to understand; this is useful for certain problems involving incidence. However the Klein model has a serious drawback: it is not conformal, i.e. the Euclidean angles in general differ from the hyperbolic angles (since orthogonal projection is not conformal).

2.10 Higher dimensional models.

We take $\mathbb{H}^n := \{(x_1, \ldots, x_n) : x_n > 0\} \subset \mathbb{R}^n$, with the metric

$$ds^2 := \frac{1}{x_n^2}(dx_1^2 + \cdots + dx_n^2).$$

Through any two points $w, z \in \mathbb{H}^n$, there is a 2-plane through w, z that meets the boundary space $\{x_n = 0\}$ orthogonally (meaning that the normal vector to the boundary is in the direction of the 2-plane). The induced geometry on this plane $P_{w,z}$ behaves exactly like 2-dimensional hyperbolic geometry. If a geodesic arc through w, z did not lie on $P_{w,z}$, then we could project it orthogonally to $P_{w,z}$ to get a strictly shorter path, contradiction.

We also have a ball model $B^n := \{x : \|x\| < 1\} \subset \mathbb{R}^n$, where $\|\cdot\|$ is the usual Euclidean norm, and

$$ds^2 := \frac{1}{(1 - \|x\|^2)^2}(dx_1^2 + \cdots + dx_n^2).$$

3 Hyperboloid model

Review of some spherical geometry properties. The round sphere metric S^n is most naturally defined by embedding S^n into $\mathbb{R}^{n+1}$ in the standard way (as locus where $x_1^2 + \cdots + x_{n+1}^2 = 1$), and taking the *induced* Riemannian metric. (Note that this does not give the same distance function on pairs of points as the restriction of the distance function from $\mathbb{R}^{n+1}$ to S^n; the distances in the former are generally larger, since one cannot short-cut through the ball.) Isometries of S^n are exactly the restrictions of isometries of $\mathbb{R}^{n+1}$ that fix the origin, i.e. orthogonal transformations. The geodesics in S^n are intersections of S^n with planes in $\mathbb{R}^{n+1}$ passing through the origin. This can be seen by using that for each such plane, there is a reflection isometry of $\mathbb{R}^{n+1}$ fixing that plane, together with Proposition 3.2 below.

We want an analogous way of thinking about $\mathbb{H}^n$ as embedded in some nice space, such that geodesics and isometries are easy to see in terms of the ambient space. The ambient space will be homeomorphic to $\mathbb{R}^{n+1}$, as for the sphere, but we will endow it with the *Minkowski metric*.

3.1 Geometry of special relativity

Postulates:

1. Any "inertial frames" (coordinate systems moving at constant relative velocity to one another) is just as good as any other for doing physics. This postulate holds in the classical physics of Galileo and Newton.

2. The speed of light c (which in math we take to be 1) is independent of frame, i.e. light looks like it's going the same speed no matter where it's emitted from. This is an empirical fact.

Einstein's train thought experiment: Suppose you are standing on a platform, and a train is moving at a (fast) constant speed relative to you. Someone on board the train has a clock. How does this clock appear to you? To analyze this we should imagine that the clock is actually just two parallel mirrors, with light bouncing back and forth between the mirrors; time is measured by recording the number of bounces. By postulate 1, this clock behaves like any other clock, say a mechanical clock, from the perspective of person on the train; otherwise the person could distinguish her reference frame from another.

Now from the perspective of the platform, the light takes a longer path, since it must move in the direction of the train as well. By postulate 2, the speed of light is still 1. Thus the observer on the platform and the person on the train will not agree on time intervals.

Minkowski metric. We model n-dimensional space, together with one dimension worth of time as *spacetime* $\mathbb{R}^{n,1} = \{(x_0, \dots, x_n) : x_i \in \mathbb{R}\}$, but with quadratic form

$$Q(x) = \|x\|^2_{n,1} := -x_0^2 + x_1^2 + \cdots + x_n^2.$$

An *event* happens at a particular point in spacetime, i.e. it has both spacial coordinates and time coordinate (x_0) with respect to the intertial frame defining the coordinate system.

We can think of our physical universe as $\mathbb{R}^{3,1}$, at least locally.

Notice that Q above is *indefinite*: a non-zero vector can have positive, zero, or negative value of Q. This means that Q does not give us a distance in the way we usually think of it (always non-negative). Nevertheless Q has physical/geometric significance.

A vector x (which we think of as a difference between two events, one happening at x and the other at 0, the origin of our coordinate system) falls into one of the three categories:

1. *Time-like* when $Q(x) < 0$. These correspond to directions in spacetime that are achievable by some frame F moving at less than the speed of light. In this case, we can interpret $\sqrt{-Q(x)}$ as the *elapsed time* measured by the observer moving in frame F.

2. *Light-like* when $Q(x) = 0$. These correspond to directions in spacetime along which light travels.

3. *Space-like* when $Q(x) > 0$. These events are too far apart in space relative to their separation in time for an observer to move from one to another. However, in this case, one can find a moving frame F in which the two events occur *simultaneously*. We interpret $\sqrt{Q(x)}$ as the *space* distance between these two events measured in frame F.

One can show that the *spacetime interval* $Q(w - z)$ between two events w, z is invariant under any change of coordinates coming from a change of inertial reference frame (which are called Lorentz transformations). Thus it is a natural invariant "distance" in the geometry of special relativity.

3.2 Hyperboloid model in Minkowski space

The hyperboloid model will be a component of the space

$$H := \{x \in \mathbb{R}^{n,1} : Q(x) = -1\},$$

with the induced metric from Q on $\mathbb{R}^{n,1}$.

Proposition 3.1. *The induced metric on H is* positive definite*, aka Riemannian (in contrast to Q on $\mathbb{R}^{n,1}$ itself).*

Proof. To see this, first we need to understand the tangent space T_xH to H at some $x \in H$. Since H is cut out by a smooth function Q, this can be thought of as the kernel of the derivative DQ_x (provided this derivative map is surjective to $\mathbb{R}$, which will follow from the next calculation). We compute

$$DQ_x = (-2x_0, 2x_1, \ldots, 2x_n).$$

Hence

$$T_xH = \{y : \langle x, y \rangle = 0\},$$

where $\langle , \rangle$ is the inner product associated to Q i.e.

$$\langle (w_0, \ldots, w_n), (v_0, \ldots, v_n) \rangle = -w_0v_0 + w_1v_1 + \cdots + w_nv_n.$$

So T_xH is the orthogonal complement of y with respect to this inner product.

Now we want to show that Q restricted to T_xH is positive definite. The idea is that x already "uses up a dimension of negative length vectors", so the directions in the complement must be positive, since Q has signature $(n, 1)$. More formally, suppose there exists some $y \neq 0$ with $\langle y, x \rangle = 0$ and $\langle y, y \rangle \leq 0$. Then every vector in $\operatorname{span}\{x, y\}$ has non-positive Q-norm, since both x, y have this property, and $\langle x, y \rangle = 0$. Since $\langle x, x \rangle = -1 \neq 0$, x and y cannot be linearly dependent, hence $\operatorname{span}\{x, y\}$ has dimension 2. On the other hand, we can find a subspace $P \subset \mathbb{R}^{n,1}$ of dimension n on which the restriction of $\langle , \rangle$ is positive definite (spanned by e.g. $(0, 1, 0, \ldots, 0), (0, 0, 1, 0, \ldots, 0), \ldots$). So by dimension count, $\operatorname{span}\{x, y\}$ and P must intersect non-trivially; a non-zero vector w in their intersection satisfies $Q(w) \leq 0$ and $Q(w) > 0$, contradiction. ∎

Hyperboloid model. We observe that H is a hyperboloid; in particular it has two sheets (corresponding to x_0 positive or negative). Each is topologically an open disc. The *hyperboloid model* of hyperbolic geometry is one sheet of this, say

$$H^+ := \{x = (x_0, \ldots, x_n) \in \mathbb{R}^{n,1} : Q(x) = -1, x_0 > 0\},$$

with the Riemannian metric defined as above. (We can also define the model as the projectivization of H.)

We would like to show that this space is isometric to our previous models of hyperbolic space, but first we will explore isometries and geodesics.

Isometries. Any bijection of $\mathbb{R}^{n,1}$ preserving $\langle , \rangle$ restricts to an isometry of H. Such maps must be linear, and so they form a matrix group, denoted $O(n, 1)$. The index two subgroup of $O(n, 1)$ preserving H^+ is denoted $O^+(n, 1)$ and is known as the *Lorentz group*; it is exactly the isometry group of H^+. The orientation preserving isometry group of H^+ is the subgroup of $O^+(n, 1)$ of matrices of determinant 1, i.e. $SO^+(n, 1)$. To prove these facts, use that isometries are characterized by action on a frame (as in HW01), and that $O^+(n, 1)$ acts transitively on points in H^+ (possible future HW exercise).

Geodesics.

Proposition 3.2. *Let M be a Riemannian manifold, and $\phi : M \to M$ an isometry. Then the fixed point set of S is a totally geodesic submanifold (i.e. any geodesic in S with respect to the induced metric is also a geodesic in M).*

Proof. (We will take it for granted that S is a smooth submanifold.) Let $p \in S$, $v \in T_pS$, and γ the unique geodesic, wrt the induced metric on S, that is tangent to v at p (we use the fact that there exists exactly one such geodesic). It suffices to show that γ is also a geodesic wrt the M metric. Let γ' be the M-geodesic tangent to v. Note that $\phi(\gamma') = \gamma'$, since $\phi(\gamma')$ is an M-geodesic tangent to $D\phi(v) = v$ (since ϕ is the identity on S). We then also easily see that ϕ is the identity on γ'. But this means $\gamma' \subset S$. An M-geodesic that lies in S is also an S-geodesic. Thus $\gamma = \gamma'$ since they're both S-geodesics tangent to v. ∎

Proposition 3.3. *Geodesics of H^+ are intersections of 2-planes S through the origin in $\mathbb{R}^{n,1}$ with H^+ (assuming that the plane intersects H^+).*

Proof. Observe that any such S intersects H^+ in a 1-dimensional curve, let p be some point on it, and let v be the tangent vector to the curve at p. Note that S is spanned by p and v. Now suppose that p is the point $(1, 0, \ldots, 0)$. The tangent plane T_pH^+ is the horizontal plane, and v lies in this. We can find a reflection isometry $\phi \in O^+(n, 1)$ about the plane S (by using a block diagonal matrix with a 1 in the upper left, and $O(n)$ matrix in lower right); the fixed point set is exactly S. Since ϕ is an isometry of H^+, by Proposition 3.2, $S \cap H^+$ is a geodesic.

For general p, we reduce to the above using transitivity of $O(n, 1)$, applying an isometry taking p to $(1, 0, \ldots, 0)$ (see Homework 2).

For any point $p \in H^+$ and $v \in T_pH^+$, we can produce a plane S through p and 0 tangent to v. The resulting geodesic is tangent to v at p. Since geodesics are uniquely determined by first order behavior at a point, we have produced all the geodesics. ∎

3.3 Map from hyperboloid to disc

There is a nice map $p : \mathbb{H}^+ \to K$ from the hyperboloid model to Klein model. We can think of this concretely as projection of H^+ to the disc $\{x_0 = 1, x_1^2 + \cdots + x_n^2 = 1\}$, from the origin. We can also think of this in terms of the projectivization map $\mathbb{R}^{n+1} \to RP^n$.

Proposition 3.4. *The projection p from hyperboloid model to Klein model is an isometry.*

Proof. We see immediately from the description of geodesics in the two models that p takes geodesics to geodesics. The hyperbolic metric can be recovered, up to scale, just from knowledge of the geodesics (this is not obvious and a bit subtle; also one needs to assume that the metric is complete). Since the pushforward of the metric on H^+ by the map p has the same geodesics as the metric from the Klein model on the disc, the two metrics agree up to scale, and it turns out the scale factor is actually 1.

∎

Composing the map p with the map $K \to D$, the Poincaré disc, discussed in Section 2.9, we get an isometry $H^+ \to D$.

4 Computation of hyperbolic area

4.1 Hyperbolic triangles

We begin by computing the area of an *ideal triangle* in $\mathbb{H}$, i.e. all the vertices are on the boundary $\partial\mathbb{H}$. We will use that all such triangles are congruent (HW exercise). So it suffices to work with the triangle T in the upper-half space with vertices at $\infty, -1, 1$. Then we compute

$$\begin{aligned}\text{area}(T) &= \int_{-1}^{1}\int_{\sqrt{1-x^2}}^{\infty}\frac{1}{y^2}dxdy = \int_{-1}^{1} -y^{-1}\Big|_{\sqrt{1-x^2}}^{\infty}dx \\ &= \int_{-1}^{1}\frac{1}{\sqrt{1-x^2}}dx = \int_{-\pi/2}^{\pi/2}\frac{1}{\sqrt{1-\sin^2\theta}}\cos\theta d\theta \\ &= \int_{-\pi/2}^{\pi/2} d\theta = \pi.\end{aligned}$$

Theorem 4.1. *The area of a hyperbolic triangle ABC with angles $\theta_1, \theta_2, \theta_3$ is*

$$\pi - (\theta_1 + \theta_2 + \theta_3).$$

Proof. All hyperbolic triangles with angles $\theta, 0, 0$ are congruent (HW exercise); let $A(\theta)$ be the area of such a triangle. First we show that $S(\theta) := \pi - A(\theta)$ is additive, i.e. $S(\theta_1 + \theta_2) = S(\theta_1) + S(\theta_2)$. Let ABC, ACD be two triangles, with angles $\angle BAC = \theta_1$ and $\angle CAD = \theta_2$, and all other angles 0. Then $ABC \cup ACD = BCD \cup ABD$, and the two triangles on each side of the equality are disjoint from one another. Hence $A(\theta_1) + A(\theta_2) = \pi + A(\theta_1 + \theta_2)$. This implies the desired additivity property of S.

Note that S is also continuous; together with additivity, this implies linearity. We can show directly that $S(\pi) = \pi$ (limiting argument), and so we find that $S(\theta) = \theta$ for all θ, and hence $A(\theta) = \pi - \theta$.

To deal with an arbitrary triangle ABC, we consider the hyperbolic rays $\overrightarrow{AB}, \overrightarrow{BC}, \overrightarrow{CA}$, which meet the ideal boundary at A', B', C', respectively. Then $A'B'C' = ABC \cup A'BB' \cup B'CC' \cup C'AA'$. The triangles on the right are disjoint, so, using the above, we get

$$\pi = \text{area}(A'B'C') = \text{area}(ABC) + \theta_2 + \theta_3 + \theta_1$$

This implies the desired result. ∎

As a consequence of the above, the angle sum of a hyperbolic triangle is always strictly less than π (it can be proved in neutral geometry, i.e. geometry without the parallel postulate, that the angle sum of a triangle is always at most π).

Small triangles with respect to any Riemannian metric (in particular, hyperbolic metric) are nearly Euclidean. This is consistent with the above, since the angle sum of such triangles is close to π.

Proposition 4.2. *Similar hyperbolic triangles are congruent.*

Proof. Let the two similar triangles be ABC and $A'B'C'$. Since the case of triangles with three or two ideal vertices is handled in the homework, we can assume that $A, B \in \mathbb{H}$. Since isometries act transitively on tangent directions, we can assume that $A = A'$, and that B, B' lie on the same ray through A, as do C, C'. If $B = B'$ and $C = C'$, then the two triangles are clearly congruent. So suppose, for contradiction, that WLOG B' is farther from A than B.

Now the geodesics $\overline{BC}$ and $\overline{B'C'}$ cannot meet in $\mathbb{H}$; if they intersected at some D, then $BB'D$ would be a triangle with angle sum at least π, contradicting Theorem 4.1. This implies that triangle ABC is contained within $AB'C'$, and in fact the complement of the former in the latter has positive area. But this contradicts Theorem 4.1, since similar triangles have the same (finite) area. ■

4.2 Discs and circles

Recall that the hyperbolic trig functions sinh, cosh parametrize points on the hyperbola $x^2 - y^2 = 1$, i.e. $\cosh^2 x - \sinh^2 x = 1$. One can also think of them as solving the system of differential equations $f' = g, g' = f$. Explicitly

$$\cosh x = \frac{e^x + e^{-x}}{2},$$
$$\sinh x = \frac{e^x - e^{-x}}{2}.$$

We start with a basic computation.

Proposition 4.3. *In the disc model $\mathbb{D}$, the hyperbolic distance between 0 and z satisfies*

$$d_{\mathbb{D}}(0, z) = 2\tanh^{-1}|z|.$$

Proof. By applying an isometry, we can assume that $z = R$ is on the positive real line. Then using the expression for the hyperbolic metric we get:

$$d_{\mathbb{D}}(0, R) = \int_0^R \frac{2}{1 - x^2} dx = 2\tanh^{-1} R.$$

■

Proposition 4.4. *The circumference of the hyperbolic circe $S(x, r)$ equals $2\pi \sinh r$.*

Proof. We can work in the disc $\mathbb{D}$, and since all circles of radius r are congruent, we can assume that $x = 0$. By Proposition 4.3, this hyperbolic circle is just the Euclidean circle of Euclidean radius $R = \tanh(r/2)$. The Euclidean circumference of this is $2\pi \tanh(r/2)$. The hyperbolic metric along this circle is just the Euclidean metric rescaled by the factor $2/(1-R^2) = 2/(1-\tanh^2(r/2))$. Hence the hyperbolic length is

$$2\pi \tanh(r/2) \cdot \frac{2}{1-\tanh^2(r/2)} = 4\pi \sinh(r/2)\cosh(r/2) = 2\pi \sinh r.$$

∎

Proposition 4.5. *The area of the hyperbolic ball $B(x, r)$ satisfies*

$$\text{area}(B(x, r)) = 4\pi \sinh^2(r/2).$$

Proof. By applying an isometry, we can assume that the ball is centered at 0 in the disc model. By Proposition 4.3, we know that this ball corresponds to $\{z : |z| < R\} \subset \mathbb{D}$, where $R = \tanh(r/2)$. Using the expression for the metric in $\mathbb{D}$, and integrating with polar coordinates with the change of variables $u = 1 - r^2$, we get

$$\begin{aligned}
\text{area}(B(0, r)) &= \int_{|z|<R} \frac{4}{(1-|z|^2)^2} dxdy = \int_0^{2\pi} \int_0^R \frac{4}{(1-r^2)^2} r dr d\theta \\
&= \int_0^{2\pi} \int_1^{1-R^2} -2u^{-2} du d\theta = \int_0^{2\pi} 2\left(\frac{1}{1-R^2} - 1\right) d\theta \\
&= 4\pi \frac{R^2}{1-R^2} = 4\pi \frac{\tanh^2(r/2)}{1-\tanh^2(r/2)} = 4\pi \sinh^2(r/2).
\end{aligned}$$

∎

Note that the area and circumference of the balls grows *exponentially* as a function of the radius.

5 Classification of hyperbolic isometries

5.1 Algebraic approach in dimension 2

We begin with the conjugacy problem for $SL_2(\mathbb{R})$, which is closely connected to isometries of $\mathbb{H}^2$.

Theorem 5.1. *Every $A \in SL_2(\mathbb{R})$ is conjugate (in $SL_2(\mathbb{R})$) to exactly one of*

(i) $\begin{pmatrix} \lambda & 0 \\ 0 & 1/\lambda \end{pmatrix}$ *for some $\lambda \in \mathbb{R}$ with $|\lambda| > 1$, if $|\operatorname{tr} A| > 2$.*

(ii) $\pm\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \pm\begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}$, *or* $\pm\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, *if $|\operatorname{tr} A| = 2$.*

(iii) $\begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$, *for some* $\theta \in (-\pi, 0) \cup (0, \pi)$, *if* $|\operatorname{tr} A| < 2$.

Proof. We begin by considering the Jordan canonical form of A, which (also use determinant 1 condition) must be one of

$$\begin{pmatrix} \lambda & 1 \\ 0 & 1/\lambda \end{pmatrix}, \pm \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix},$$

for some $\lambda \in \mathbb{C}^*$. That is, A is conjugate in $GL_2(\mathbb{C})$ to some B in the above list.

If B is of the second type, or one of the first type with $\lambda \in \mathbb{R}$, then since such B has real entries, we can conclude that A is conjugate to this B in $GL_2(\mathbb{R})$ (since if real matrices are conjugate over $\mathbb{C}$, they are in fact conjugate over $\mathbb{R}$; this follows from rational canonical form).

Now suppose B is of the first type with $\lambda \notin \mathbb{R}$. Such matrices are determined by their eigenvalues. The matrices of type (iii) are also conjugate to matrices of the same type, and these matrices run over all possible pairs of non-real eigenvalues. Thus B is conjugate to a matrix of type (iii). As in the previous case, we can take the conjugating matrix to be in $GL_2(\mathbb{R})$.

Thus we have found a matrix QAQ^{-1} of the desired form from the theorem statement, with $Q \in GL_2(\mathbb{R})$, but $\det Q$ is not guaranteed to be 1. By multiplying Q by a real scalar, we can arrange that $\det Q = \pm 1$.

For type (i), if $\det Q = -1$, we can replace Q by $Q \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, which does have determinant 1. (And then we can further conjugate by $\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ if necessary to switch the diagonal entries).

For type (ii), if $\det Q = -1$, we can replace Q by $Q \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$, which has determinant 1. This potentially introduces a negative sign in the upper-right hand corner, which is unavoidable.

For type (iii), if $\det Q = -1$, we again replace Q by $Q \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. The matrix we get out is $\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} = \begin{pmatrix} \cos(-\theta) & \sin(-\theta) \\ -\sin(-\theta) & \cos(-\theta) \end{pmatrix}$, so it is still of the desired form.

■

Corollary 5.2. *Every $A \in PSL_2(\mathbb{R})$ is conjugate (in $PSL_2(\mathbb{R})$) to exactly one of the classes represented by*

(i) $\begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$*, for some $\theta \in (0, \pi)$, if $|\operatorname{tr} A| < 2$.*

(ii) $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}$*, or* $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$*, if $|\operatorname{tr} A| = 2$.*

(iii) $\begin{pmatrix} \lambda & 0 \\ 0 & 1/\lambda \end{pmatrix}$ *for some $\lambda \in \mathbb{R}$ with $\lambda > 1$, if $|\operatorname{tr} A| > 2$.*

(Note that $|\operatorname{tr} A|$ is well-defined for elements of $PSL_2(\mathbb{R})$.)

The first two types above can be understood well in the upper-half space model. Type (iii) ("hyperbolic") corresponds to $z \mapsto \lambda^2 z$, while Type (ii) corresponds to the identity map, or $z \mapsto z \pm 1$ ("parabolic"). Type (i) ("elliptic"), which as an isometry of the upper-half space fixes i, is best understood by moving to the Poincaré disc model; there it corresponds to rotation about the origin (when i is mapped to 0).

5.2 Geometric approach

Proposition 5.3. *Let $\phi \in \operatorname{Isom}(\mathbb{H}^n)$. Then either*

(i) ϕ has at least one fixed point in $\mathbb{H}^n$ ("elliptic")

(ii) ϕ has no fixed points in $\mathbb{H}^n$, and exactly one fixed point in $\partial\mathbb{H}^n$ ("parabolic"),

(iii) ϕ has no fixed points in $\mathbb{H}^n$, and exactly two fixed points in $\partial\mathbb{H}^n$ ("loxodromic").

Proof. Since $\overline{\mathbb{H}^n}$ is homeomorphic to a closed ball, and ϕ is continuous, ϕ has at least one fixed point in $\overline{\mathbb{H}^n}$ by the Brouwer fixed point theorem. So to prove the proposition, it suffices to show that if ϕ has three or more fixed points in $\partial\mathbb{H}^n$, then it has a fixed point in $\mathbb{H}^n$. Consider the geodesic γ joining two of the fixed points, and let α be the geodesic perpendicular to γ and limiting to the other fixed point. We see the intersection point of α, γ is fixed by ϕ, but this lies inside $\mathbb{H}^n$, contradiction. ■

Horocycles and horospheres. Given a point $p \in \partial\mathbb{H}^n$, we can consider all the geodesics that end at p. A *horosphere centered at* p is a hypersurface that is perpendicular to all these geodesics. In the Poincaré ball B^n, these are spheres tangent to the boundary sphere ∂B^n at p. In the upper-half space model, horospheres centered at ∞ are horizontal planes. From this description, we see that the intrinsic geometry of each horosphere is Euclidean. The horospheres are *not* totally geodesic.

Note that if ϕ is an isometry with p a fixed point, then ϕ maps a horocycle centered at p to another horocycle centered at p.

Proposition 5.4. *Let* $\phi \in \mathrm{Isom}(\mathbb{H}^n)$*. Then*

(i) If ϕ *is elliptic, then, in* B^n *with* 0 *a fixed point, we have* $\phi(v) = Av$ *for some* $A \in O(n)$*.*

(ii) If ϕ *is parabolic, then in the upper-half space* $\mathbb{H}^n$ *with* ∞ *as the fixed point,* $\phi(x, x_n) = (Ax + b, x_n)$*, for* $A \in O(n-1)$ *and* $b \in \mathbb{R}^{n-1}$ *(here* (x, x_n) *is short-hand for* $(x_1, \ldots, x_n)$*).*

(iii) If ϕ *is loxodromic, then in the upper-half space model* $\mathbb{H}^n$ *with* $0, \infty$ *as fixed points,* $\phi(x, x_n) = \lambda(Ax, x_n)$*, for some* $A \in O(n-1)$*, and* λ *positive real with* $\lambda \neq 1$*.*

Proof. For (i), we can find such an A that gives an isometry that agrees with ϕ in its action on an orthonormal frame at 0. Since an isometry is determined by this data, $\phi(x) = Ax$.

For (ii), begin by noting that since ϕ takes each geodesic through ∞ to another geodesic through ∞, it also takes each horosphere centered at ∞ to a horosphere centered at ∞. Horospheres can be parametrized by height (x_n coordinate in upper-half space). So ϕ induces a map on $\mathbb{R}^+$.

We claim that this map is the identity, i.e. each horosphere centered at ∞ maps to itself under ϕ. Otherwise, we assume that there is some horosphere S_t that maps to $S_{t'}$ with $t' < t$ (otherwise, replace ϕ by ϕ^{-1}. Now the map $f : S_{t'} \to S_t$ given by $(x_1, \ldots, x_{n-1}, t') \mapsto (x_1, \ldots, x_{n-1}, t)$ is a strict contraction. Since ϕ is an isometry, the map $g := f \circ \phi|_{S_t} : S_t \to S_t$ is also a strict contraction. The Banach fixed point theorem then implies that g has a fixed point. This means that ϕ takes some point $p \in \mathbb{H}^n$ to a point on the same geodesic γ through ∞. Then γ must fix the other endpoint of γ on the boundary, but then it cannot be parabolic.

Now on each such horosphere S_t, ϕ must act as a Euclidean isometry, i.e $(x, t) \mapsto (Ax + b, t)$, where $A \in O(n)$. We see that A, b do not depend on t, since vertical geodesics must be mapped to vertical geodesics.

For (iii), note first that the vertical geodesic γ connecting the two fixed points $0, \infty$ must be preserved by ϕ, and ϕ acts by a translation on γ. Let λ be such that $(0, \ldots, 0, 1) \mapsto (0, \ldots, 0, \lambda$. As in the parabolic case, ϕ takes each horizontal horosphere to a horizontal horosphere. If we compose ϕ with the map $v \mapsto \lambda^{-1} v$, we get a map that preserves each horizontal horosphere, and is the identity on the line through $0, \infty$. It follows that the composition must be given by $(x, x_n) \mapsto (Ax, x_n)$. Then we get that ϕ has the desired form.

■

Isometries of $\mathbb{H}^3$**.** A loxodromic isometry of $\mathbb{H}^3$ has a "translation" component (given by λ), as well as a rotation component (given by $A \in O(2)$).

There is an alternate description of $\mathrm{Isom}^+(\mathbb{H}^3)$.

Proposition 5.5. *We have* $\mathrm{Isom}^+(\mathbb{H}^3) \cong PSL_2(\mathbb{C})$*.*

Proof. This is seen using the Poincaré ball model. The group of isometries (including orientation reversing) is generated by hyperbolic reflections, i.e. inversions through spheres orthogonal to the boundary sphere S^2. Such an inversion restricts to an inversion on S^2, which we think of as the Riemann sphere. And conversely, any inversion on the Riemann sphere extends to a hyperbolic reflection of the ball. It follows that $\mathrm{Isom}(\mathbb{H}^3)$ is the group generated by Mobius transformations together with a single inversion. And taking the subgroups of orientation-preserving transformations gives the desired result. ∎

6 Hyperbolic manifolds

Definition 6.1. A hyperbolic n-manifold M is a Riemannian manifold that is locally isometric to $\mathbb{H}^n$, i.e. any $x \in M$ has a neighborhood U such that U is isometric to an open subset of $\mathbb{H}^n$.

Bolza surface. In HW02, a regular hyperbolic octagon with all angles $\pi/4$ was constructed. By "gluing" together the opposite sides using loxodromic isometries, we define a Riemannian metric on the quotient of the octagon by identification of opposing sides. The resulting X is a hyperbolic surface of genus 2, called the Bolza surface. The construction can be varied by e.g. changing some side lengths.

Hyperbolic annulus. Fix $\lambda \in \mathbb{R}_{>0}$. Then the quotient $A_\lambda = \mathbb{H}/\langle z \mapsto \lambda z\rangle$ is an annulus, which inherits a hyperbolic structure.

6.1 Covering spaces

Reference for this material: Hatcher "Algebraic Topology" Ch. 1.3.

Definition 6.2. If $\tilde{X}, X$ are topological spaces, a continuous map $f : \tilde{X} \to X$ is a *covering map* if every $x \in X$ has a neighborhood N that's evenly covered, i.e. $f^{-1}(N) = \coprod_\alpha M_\alpha$, where M_α are open, and $f|_{M_\alpha} : M_\alpha \to N$ is a homeomorphism.

Example: For an integer n, the map $S^1 \to S^1$ (thought of as unit circle in $\mathbb{C}$) given by $z \mapsto z^n$ is a covering map.

Theorem 6.3. *Every connected manifold X (or more generally, space that is path-connected, locally path-connected, and semi-locally simply connected) has a* universal cover*, i.e. a covering $f : \hat{X} \to X$ with $\hat{X}$ simply connected.*

Proof sketch. We define $\hat{X}$ to be the space of paths in X starting at some fixed basepoint x_0, up to homotopy relative to the endpoints. ∎

Example: $\mathbb{R} \to S^1$ given by $x \mapsto e^{ix}$ is a universal covering.

A continuous map $f : (X', x'_0) \to (X, x_0)$ of pointed topological spaces (which means $f(x'_0) = x_0$) induces a homomorphism of fundamental groups $f_* : \pi_1(X', x'_0) \to \pi_1(X, x_0)$.

If $f_1 : \tilde{X}_1 \to X, f_2 : \tilde{X}_2 \to X$ are covering maps, an isomorphism between them is a homeomorphism $h : \tilde{X}_1 \to \tilde{X}_2$ such that $f_2 \circ h = f_1$. These are also called *deck transformations.* There is a similar notion of base-point preserving isomorphism.

Theorem 6.4. *Given (X, x_0), there is a bijection between basepoint-preserving isomorphism classes of covering spaces of X and subgroups of $\pi_1(X, x_0)$. It is given by taking $p : (\tilde{X}, \tilde{x}_0) \to (X, x_0)$ to $p_*(\pi_1(X, x_0))$.*

6.2 Universal cover of hyperbolic manifold

Given a hyperbolic manifold M, we can form the topological universal cover $\hat{M}$. This space inherits a hyperbolic manifold structure: locally lift the metric from M. The below tells us what this $\hat{M}$ has to be.

Proposition 6.5. *Every simply connected, complete hyperbolic n-manifold M is isometric to $\mathbb{H}^n$.*

Proof. Choose some $x \in M$, and an open neighborhood U of x with map $f : U \to \mathbb{H}^n$ that's an isometry to its image. We will extend f to a map $\tilde{f} : M \to \mathbb{H}^n$, called a developing map. Given some point $y \in M$, choose a path γ connecting x to y. Each point along γ has a neighborhood that is isometric to a round disc in $\mathbb{H}^n$. By compactness, we can cover γ with finitely many such discs. Using Lemma 6.6 below, we can arrange that the maps for the discs agree on consecutive overlaps. One can also see that replacing γ by a homotopic arc γ' results in the same value of $\tilde{f}(y)$. Thus we get a map $\tilde{f} : M \to \mathbb{H}^n$, and from the way it was constructed we see that it is a local isometry.

Now we use that if $f : M \to N$ is a local isometry, and M complete, then f is a covering map. Thus our $\tilde{f}$ is a covering, and since the only covering between simply connected spaces is isomorphic to the identity map (follows from Theorem 6.3), we get that $\tilde{f}$ is a homeomorphism. A homeomorphism that is locally an isometry is globally an isometry, so we're done. ∎

Lemma 6.6. *If U, V are connected open subsets of $\mathbb{H}^n$, and $f : U \to V$ an isometry, then f extends uniquely to an isometry $\mathbb{H}^n \to \mathbb{H}^n$.*

Proof. Pick an orthonormal frame at some $x \in U$. The map f is determined by its action on this frame (since for any other y, can find a concatenation of finitely many geodesic arcs connecting x to y; the action on the frame tells us what the image of this geodesic is, and since f is an isometry, we know how far along the next endpoint will be, and then we can continue for each segment). But there is also an isometry $\phi : \mathbb{H}^n \to \mathbb{H}^n$ that acts the same way on the frame. This ϕ must agree with f on U, since both are given by the same description in terms of images of geodesics. Hence ϕ is the desired extension. ∎

6.3 Quotients of hyperbolic space

An arbitrary quotient of a space by a continuous group action need not be very nice. For instance $\mathbb{R}/\mathbb{R}^*$ is not Hausdorff or even T_1: it consists of two points, $[0], [1]$, but every neighborhood of $[0]$ contains $[1]$.

Definition 6.7. A continuous group action G on X is said to be *properly discontinuous* if for any $K \subset X$ compact, the set

$$\{g : gK \cap K \neq \emptyset\}$$

is finite. The action is *free* if for all $g \neq 1$, the fixed point set $\{x \in X : g(x) = x\}$ is empty.

Examples:

- Action of a single loxodromic on $\mathbb{H}$ is free and properly discontinuous.
- Action by an elliptic is not free, and if it is infinite order, the action is not properly discontinuous. If the elliptic has order k, we get a properly discontinuous action by $\mathbb{Z}/k\mathbb{Z}$ (since every action by a finite group is trivially properly discontinuous).
- Action on $\mathbb{R}^2 - \{0\}$ by $(x, y) \mapsto (2x, y/2)$ is not properly discontinuous, though it is *wandering*: every point has a neighborhood that intersects only finitely many of its translates under the group action. The quotient by the action is not Hausdorff, because a sufficiently high forward iterate of any neighborhood of $(0, 1)$ intersect a sufficiently high backward iterate of any neighborhood of $(1, 0)$.

Proposition 6.8. *Let G act on a Hausdorff, connected space X. The following are equivalent:*

(i) The action is free and properly discontinuous.

(ii) The quotient X/G is Hausdorff, and $X \to X/G$ is a covering map.

Corollary 6.9. *Every "nice space" (path-connected, locally path-connected, semi-locally simply connected, and Hausdorff) X is the quotient $\hat{X}/G$, where $\hat{X}$ is the universal cover, and G is a group acting properly discontinuously (the deck group).*

Corollary 6.10. *Any complete hyperbolic n-manifold M is a quotient of $\mathbb{H}^n$ by group of isometries acting freely and properly discontinuously.*

Proof. By Proposition 6.5, the universal cover $\hat{M}$ is isometric to $\mathbb{H}^n$. The deck group of the universal cover acts by isometries for the pull-back metric on M. The map $\hat{M} \to M$ satisfies Proposition 6.8 (ii), hence it satisfies (i), i.e. it is free and properly discontinuous. ■

For actions of subgroups of Isom on $\mathbb{H}^n$, proper discontinuity is equivalent to *discreteness* (as a subset of Isom($\mathbb{H}^n$), with its natural topology). The obstruction to freeness for actions by subgroups of Isom on $\mathbb{H}^n$ is exactly the presence of elliptic elements. (An example of a large discrete subgroup is $PSL_2(\mathbb{Z})$, though this contains elliptics, such as $z \mapsto -1/z$.)

Hence we get a correspondence:

$$\{\text{complete hyperbolic manifolds, up to isometry}\} \leftrightarrow$$
$$\{\text{discrete subgroups of Isom}(\mathbb{H}^n) \text{ without elliptics, up to conjugation}\}.$$

The "up to conjugation" equivalence is present since we have a choice of isometry between $\hat{M}$ and $\mathbb{H}^n$; different choices lead to conjugate subgroups.

Definition 6.11. A tessellation of $\mathbb{H}^n$ is a locally finite set of polyhedra that cover $\mathbb{H}^n$ and intersect only in common faces.

Definition 6.12. A fundamental domain for the action of $\Gamma \subset \mathbb{H}^n$ is a polyhedron in $\mathbb{H}^n$ whose translates are distinct and form a tessellation of $\mathbb{H}^n$.

Examples of hyperbolic manifolds:

- $\mathbb{H}/\langle z \mapsto z+1 \rangle$
- Bolza surface. For each pair of opposite sides, there is a loxodromic isometry mapping one to the other (the axis is the common perpendicular bisector to the two sides). Let Γ be the group generated by these. Then the Bolza surface is $\mathbb{H}/\Gamma$. The octagon is a fundamental domain for the action of Γ.
- Seifert-Weber dodecahedral space: identify opposite faces of a dodecahedron via a 3/10 twist. Edges are glued in 6 groups of 5. The dihedral angle between faces of Euclidean regular dodecahedron is approximately 117 degrees. An ideal hyperbolic regular dodecahedron has 60 degree dihedral angles. By the intermediate value theorem, there is a hyperbolic regular dodecahedron with angles 72 degrees. Doing the gluing with this geometric realization givs the Seifert-Weber hyperbolic 3-manifold, which is compact. This is one of the simplest examples of a closed hyperbolic 3-manifold.

6.4 Hyperbolic orbifolds

A natural discrete subgroup of $PSL_2(\mathbb{R})$ is $PSL_2(\mathbb{Z})$. However, this contains elliptic elements, such as the Mobius transformation $z \mapsto -1/z$. So it does not actually correspond to a hyperbolic surface. However, the quotient $\mathbb{H}/PSL_2(\mathbb{Z})$ is still a fairly nice object. It is a hyperbolic *orbifold*: every point has a neighborhood modeled on a quotient of $\mathbb{H}^n$ by a finite group of isometries (if the group at every point is the trivial group, then we get an actual hyperbolic surface). The action of $PSL_2(\mathbb{Z})$ on $\mathbb{H}$ has a nice fundamental domain, the region bounded by the unit circle and the lines $\text{Re}(z) = \pm 1/2$. The points $e^{2\pi/6}, e^{2\pi/3}, i$ are orbifold points. The quotient $\mathbb{H}/PSL_2(\mathbb{Z})$ is non-compact, but has finite volume.

Congruence subgroups. Now consider the group homomorphism

$$\phi_m : PSL_2(\mathbb{Z}) \to PSL_2(\mathbb{Z}/m\mathbb{Z}),$$

given by reducing all entries modulo m. The kernel of ϕ_m, denoted $\Gamma(m)$, is subgroup of $PSL_2(\mathbb{Z})$, hence also discrete, and is finite-index and normal. They are known as *principal congruence subgroups.* For $m \geq 4$, we claim that $\Gamma(m)$ contains no elliptics. In fact, the diagonal entries a, d of a representative of such an element would need to satisfy $a + d \equiv \pm 2 \bmod m$ in order to be in $\Gamma(m)$, while on the other hand to be elliptic, $a + d = 1, 0, -1$; there are no such matrices.

We thus see that for $m \geq 4$, $\Gamma(m)$ is a discrete subgroup without elliptics, and hence corresponds to a hyperbolic surface. It is a finite degree cover of $\mathbb{H}/PSL_2(\mathbb{Z})$.

In the above, we passed to a cover to get rid of all the elliptics; in general one can always do this:

Theorem 6.13 (Selberg)**.** *Every finitely generated discrete subgroup* $\Gamma \subset \mathrm{Isom}(\mathbb{H}^n)$ *has a finite index subgroup that acts freely on* $\mathbb{H}^n$*.*

Bianchi groups. There are (many) three-dimensional versions of the above. We start by thinking of $\mathrm{Isom}^+(\mathbb{H}^3)$ as $PSL_2(\mathbb{C})$. For the simplest example, take $\Gamma := PSL_2(\mathbb{Z}[i])$. This is a subgroup, since $\mathbb{Z}[i]$ is a ring, and it is easy to see that this is discrete, so the quotient is a hyperbolic orbifold. We can replace the Gaussian integers $\mathbb{Z}[i]$ by the ring of integers $\mathcal{O}_d$ in any imaginary quadratic number field $\mathbb{Q}[\sqrt{-d}]$, where d is a positive square-free integer. (Recall that $\mathcal{O}_d$ is $\mathbb{Z}[\sqrt{-d}]$ if $d \equiv 1, 2 \bmod 4$, and $\mathbb{Z}[(-1 + \sqrt{-d})/2]$ otherwise.)

The groups $PSL_2(\mathcal{O}_d)$ are known as *Bianchi groups*, and their quotients *Bianchi orbifolds.* They are all finite volume. Their geometric properties (number of cusps, volumes, etc) are related to the number theory of $\mathbb{Q}[\sqrt{-d}]$ (class group, Dedekind zeta function, etc).

6.5 Rigidity in higher dimensions

Theorem 6.14 (Mostow rigidity)**.** *If* M_1, M_2 *are complete, closed hyperbolic* n*-manifolds with* $n \geq 3$ *and* $\pi_1(M_1) \cong \pi_1(M_2)$*, then* M_1 *and* M_2 *are isometric.*

This is a very strong form of rigidity! It was also extended by Prasad to the case where M_1, M_2 are not assumed to be compact, but rather just finite volume. One often applies it in the case where one was the stronger information that M_1, M_2 are homeomorphic.

A consequence of Mostow Rigidity is that geometric invariants of these hyperbolic manifolds are in fact topological invariants. This is useful in knot theory: hyperbolic invariants of a knot complement, such as volume or length of shortest closed geodesic, are actually topological invariants (they do not depend on the way the knot is presented). Although not all knot complements are hyperbolic, in practice most are, and the above leads to practically useful ways of distinguishing knots (see SnapPea/SnapPy software).

6.6 Closed hyperbolic surfaces

Unlike in dimension $n \geq 3$, when $n = 2$, there is a great deal of flexibility. We start by recalling the topological classification.

Theorem 6.15 (Classification of surfaces)**.** *Let S be a closed, orientable surface. Then S is homeomorphic to either*

(i) the sphere (genus $g = 0$),

(ii) the torus, ($g = 1$), or

(iii) a genus g surface, for g an integer with $g \geq 2$.

We group them into the above three categories for geometric reasons that will become apparent. The *Euler characteristic* $\chi(S)$ equals $2 - 2g$. The cases correspond to positive, zero, and negative Euler characteristics, respectively.

Theorem 6.16 (Gauss-Bonnet)**.** *For X a closed, orientable surface with a Riemannian metric,*

$$2\pi \cdot \chi(X) = \int_X K dA,$$

where K is the Gaussian curvature function.

Proof for $K \equiv -1$. Consider a triangulation of the surface by geodesic segments; let v, e, f be number of vertices, edges, faces, respectively. By Theorem 4.1, the area of each triangle T equals the angle defect $A_T = \pi - \alpha + \beta + \gamma$. We can compute the area by summing over these triangles, giving

$$\int_X K dA = -\operatorname{area}(X) = -\sum_T A_T = -\pi f + \sum \alpha_i,$$

where the last sum is over all angles, and hence equals $2\pi v$. Since we have a triangulation, $3f = 2e$, hence $\chi(X) = v - e + f = v - f/2$. So continuing with the above we get

$$\int_X K dA = -\pi f + 2\pi v = \pi(2v - f) = 2\pi \cdot \chi(X),$$

as desired. ■

The round metric on the unit sphere S^2 has constant curvature 1, the flat metric on $\mathbb{R}^2$ constant curvature 0, and the hyperbolic metric on $\mathbb{H}^2$ has constant curvature -1. We can prove Gauss-Bonnet for surfaces modeled on these geometries in a similar way to the above.

It follows from Gauss-Bonnet, that for topological surfaces, at most one of the three geometries is possible.

Towards putting hyperbolic metrics on higher genus surfaces, we now develop some building blocks.

Theorem 6.17. *Given $\ell_1, \ell_2, \ell_3 \in \mathbb{R}^3_{>0}$, there exists a unique right-angled hyperbolic hexagon $AcBaCb$ (with all sides geodesic segments) and $|a| = \ell_1, |b| = \ell_2, |c| = \ell_3$.*

For a soft proof, see Thurston-Levy Figure 4.15.

One can in fact extend the above by allowing some or all cuff lengths to be zero. These will give cusped hyperbolic surfaces when suitably glued.

Pairs of pants. If we have two isometric right-angled hyperbolic hexagons $AcBaCb, A'c'B'a'C'b'$, we can glue A to A', B to B' and C to C'. This gives a surface with boundary called a *pair of pants* P. Because we are gluing along geodesic arcs of the same length, the interior of P inherits a hyperbolic metric. There are three boundary components (e.g., one is the union of a and a'), and since two right angles come together where e.g. a, a' meet, each boundary is a closed geodesic.

Conversely, given a hyperbolic pair of pants P with geodesic boundary, we can cut into two hexagons as follows. For each pair of boundary components, consider the geodesic arc γ_i of shortest length connecting the two components. These *seams* must be perpendicular to each boundary; otherwise there would be a shorter path. Cutting along the three seams gives two right-angled hyperbolic hexagons $AcBaCb, A'c'B'a'C'b'$, where the lower-case letter sides came from cutting along the seams, and thus e.g. $|a| = |a'|$. It follows from the uniqueness part of the above theorem that the two hexagons are isometric.

From the above two paragraphs, we see that the data of a single right-angled hyperbolic hexagon is equivalent to the data of a hyperbolic pair of pants with geodesic boundary. And it also follows that for any choice of positive cuff lengths, there's a unique right-angled hyperbolic hexagon with those cuff lengths.

Gluing together pants. To form a closed hyperbolic surface X, we can glue together pairs of pants along cuffs. A pair of cuffs that are glued together should have the same length. There is a choice of *twist* when two cuffs are glued together.

The Euler characteristic of a pair of pants is -1. By additivity of Euler characteristic, since $\chi(X) = 2 - 2g$, there are $2g - 2$ pants in any pants decomposition of X (additivity is over disjoint unions, but we can still use it in our setting, since each boundary circle has Euler characteristic zero).

Given a topological pants decomposition, to specify the hyperbolic structure we need: $(2g - 2) \cdot (3/2) = 3g - 3$ length parameters, and an equal number of twist parameters. Thus the "space of hyperbolic structures" on a genus g surface has (real) dimension $6g - 6$. The cuff lengths and twist parameters give *Fenchel-Nielsen coordinates* on the space of hyperbolic structures, which is thus homeomorphic $\mathbb{R}^{6g-6}_{>0} \times \mathbb{R}^{3g-3}$. However, these are coordinates on *marked* hyperbolic structures. This is called the *Teichmüller space* $\mathcal{T}_g$, and it is not the same as the *moduli space* $\mathcal{M}_g$, which is the set of hyperbolic structures up to isometry. A marked hyperbolic surface is a hyperbolic surface X equipped with a map $S \to X$, defined up to homotopy, where S is a fixed topological surface.

There are (infinitely) many different pants decompositions of a surface (even if we consider the cuff curves up to homotopy). But there are only finitely many up to homeomorphism. The isometry type of a surface appears infinitely many times in a Fenchel-Nielsen chart.

Existence of geometric pants decompositions. A missing piece in the above is: how do we know that any hyperbolic surface has a decomposition into hyperbolic pairs of pants with geodesic boundary? We start by finding closed geodesics.

The intuition for the below is that the shortest element of a homotopy class should be a geodesic.

Proposition 6.18. *In any non-trivial homotopy class of loops on a closed hyperbolic surface X, there exists a unique closed geodesic.*

Proof. Let γ be a representative of the homotopy class, and pick a basepoint $p \in X$. So we can also think of γ as an element of $\pi_1(X, p)$. Now pick a universal covering $\pi : (\mathbb{H}, \tilde{p}) \to (X, p)$. Consider the lift $\tilde{\gamma}$ of γ to $\mathbb{H}$, going through $\tilde{p}$. Now $\tilde{\gamma}$ is periodic under the action of γ on $\mathbb{H}$. This isometry is loxodromic by Proposition 6.19, hence has an axis α. Consider the straight-line homotopy from $\tilde{\gamma}$ to α. This is equivariant for the action of γ, and hence the homotopy descends to a homotopy from γ to $\pi(\alpha)$, which is a closed geodesic.

For uniqueness, use that lifts are a bounded distance apart, hence have the same endpoints, and there's only one choice of axis with these endpoints. ■

Recall that if we fix a universal covering $\mathbb{H} \to X$, then every element of $\pi_1(X)$ acts by an isometry on $\mathbb{H}$, and X is isometric to $\mathbb{H}/\pi_1(X)$.

Proposition 6.19. *Every non-trivial element ϕ of $\pi_1(X) \subset \mathrm{Isom}^+(\mathbb{H})$ is a loxodromic isometry.*

Proof. Note that ϕ cannot be elliptic, since then its action on $\mathbb{H}$ would not be free. It remains to show that ϕ cannot be parabolic; suppose the contrary. The translation length of a parabolic (i.e. the infimum of the distance between a point and its image) is zero. So for each $\epsilon > 0$, we can find $x \in \mathbb{H}$ and an arc α connecting $x, \phi(x)$ of length $< \epsilon$. Projecting this arc to X gives a closed curve on X of length $< \epsilon$. This curve represents an element in the conjugacy class of ϕ (thought of as an element of $\pi_1(X)$), and hence is not null-homotopic. But on a compact surface there is a lower bound on the lengths of curves in any non-trivial homotopy class, contradiction.

■

6.7 Other perspectives

Fix a $g \geq 2$. There are equivalences of categories between:

- (Closed hyperbolic surfaces of genus g, isometries),
- (Compact Riemann surfaces of genus g, biholomorphic maps),
- (Smooth algebraic curves of genus g, biregular morphisms).

To see the equivalences require some deep theorems. For instance, to show that a Riemann surface admits a hyperbolic metric consistent with the conformal structure requires the Uniformization Theorem. And to embed a Riemann surface as an algebraic subset of projective space, one uses Kodaira vanishing.

There is also a version for $g = 1$, where “hyperbolic” is replaced by “flat structures of area 1.”