Seminar: "Geometry Structures on manifolds"
Hyperbolic Geometry

Xiaoman Wu

December 1st, 2015

1 Poincaré disk model

Definition 1.1. (Poincaré disk model) The hyperbolic plane $\mathbf{H}^2$ is homeomorphic to $\mathbf{R}^2$, and the Poincaré disk model, introduced by Henri Poincaré around the turn of this century, maps it onto the open unit disk $\mathbf{D}$ in the Euclidean plane.

Hyperbolic straight lines, or geodesics, appear in this model as arcs of circles orthogonal to the boundary $\partial\mathbf{D}$ of $\mathbf{D}$, and every arc is one special case: any diameter of the disk is a limit of circles orthogonal to $\partial\mathbf{D}$ and it is also a hyperbolic straight line.

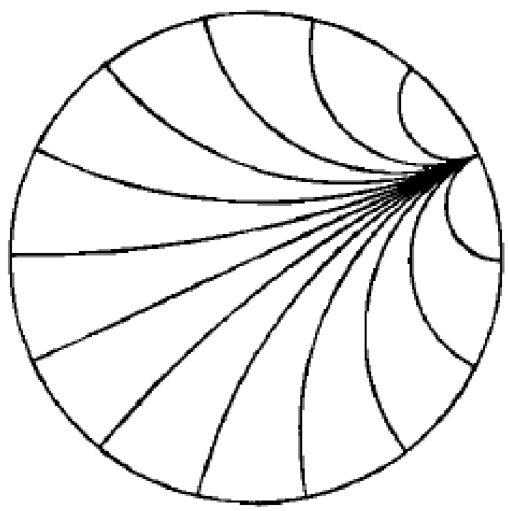

Figure 1: Straight lines in the Poincaré disk model appear as arcs orthogonal to the boundary of the disk or, as a special case, as diameters.

Define the Riemannian metric by means of this construction (Figure 2). To find the length of a tangent vector v at a point x, draw the line L orthogonal to v through x, and the equidistant circle C through the tip. The length of v (for v small) is roughly the hyperbolic distance between C and L, which in turn is roughly equal to the Euclidean angle between C and L where they meet. If we want to exact value, we consider the angle α_t of the banana built on tv, for t approaching zero: the length of v is then $d\alpha_t/dt$ at $t = 0$.

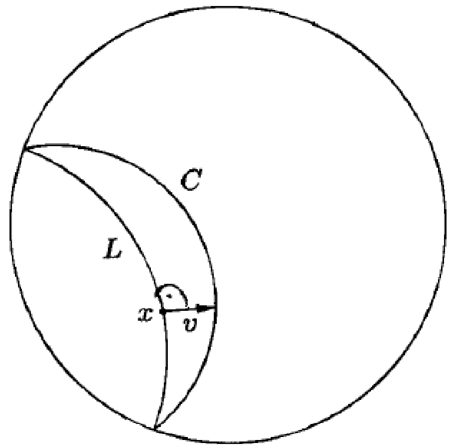

Figure 2: Hyperbolic versus Euclidean length. The hyperbolic and Euclidean lengths of a vector in the Poincaré model are related by a constant that depends only on how far the vector's basepoint is from the origin.

Setting the hyperbolic length of v equal to the banana angle in the limit when both go to 0-gives the following formula for the hyperbolic metric ds^2 as a function of the Euclidean metric dx^2:

$$ds^2 = \frac{4}{(1-r^2)} dx^2. \tag{1.1}$$

Definition 1.2. (Visual sphere) Think of an observer as a point somewhere in an n-dimensional space, with light rays approaching this point along geodesics. Each of these geodesics determines a tangent vector at the point, and the $(n-1)$-sphere of tangent vectors called the *visual sphere.*

2 The Inversive Models

Definition 2.1. (Inversion in a sphere) If $\mathbf{S} \in \mathbf{E}^n$ is an $(n-1)$-sphere in Euclidean space, the *inversion* i_S in $\mathbf{S}$ is the unique map from the complement of the center of $\mathbf{S}$ into itself that fixes every point of $\mathbf{S}$, exchanges the interior and exterior of $\mathbf{S}$ and takes spheres orthogonal to $\mathbf{S}$ to themselves.

As in the two-dimensional case, the image $i_S(P)$ of a point P in a circle $\mathbf{S}$ with center O and radius r is the point on the ray $\overrightarrow{OP}$ such that $OP \cdot OP' = r^2$.

It is somewhat annoying that inversion in a sphere in $\mathbf{E}^n$ does not map its center anywhere. We can remedy this by considering the one-point compactification $\widehat{\mathbf{E}^n} = \mathbf{E}^n \bigcup \{\infty\}$ of $\mathbf{E}^n$, which is homeomorphic to the sphere $\mathbf{S}^n$. An inversion i_S can then be extended to map the center of $\mathbf{S}$ to ∞ and vice versa, so it becomes a homeomorphism of $\widehat{\mathbf{E}^n}$.

Property 2.1. (Properties of inversions) Let $\mathbf{S}$ be $(n-1)$-dimensional proper sphere in $\mathbf{E}^n$. Then the inversion i_S is conformal, and takes spheres (of any dimension) to spheres.

Proof. For conformality, notice that any two vectors based at a point are the normal vectors to two $(n-1)$-spheres orthogonal to $\mathbf{S}$, so both the angle between them and the angle between their images equal the dihedral angle between the spheres.

The second statement follows from the plane case for spheres of codimension one by considering the symmetries around the line joining the centers of the inverted and inverting spheres; and for lower-dimensional spheres because they are intersections of spheres of codimension one. □

The *Poincaré ball model* of hyperbolic space is what we get by taking the unit ball $\mathbf{D}^n$ in $\mathbf{E}^n$ and declaring to be hyperbolic geodesics all those arcs of circles orthogonal to the boundary of $\mathbf{D}^n$. We also declare that inversions in $(n-1)$-spheres orthogonal to $\partial \mathbf{D}^n$ are hyperbolic isometries, which we will call *hyperbolic reflections*.

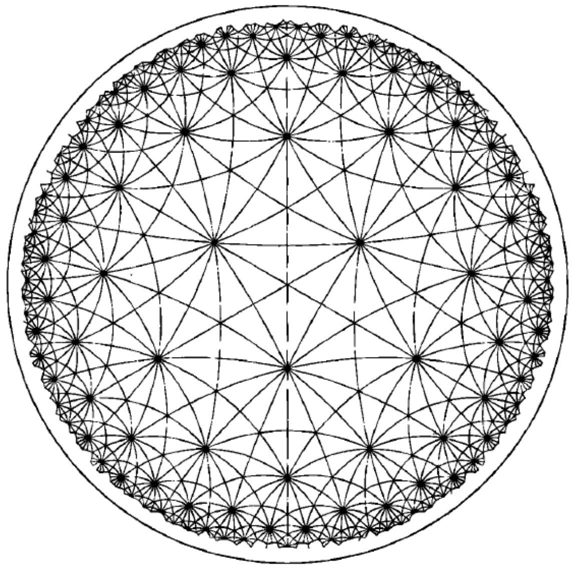

Figure 3: Hyperbolic tiling by 2-3-7 triangles. The hyperbolic plane laid out in congruent tracts, as seen in the Poincaré model. The tracts are triangles with angle $\pi/2$, $\pi/3$, $\pi/7$. Courtesy HWG Homestead Bureau.

We see that distances are greatly distorted in the Poincaré Model: the Euclidean image of an object has size roughly proportional to its Euclidean distance from the boundary $\partial \mathbf{D}^n$, if this distance is small (Figure 3). A person moving towards $\partial \mathbf{D}^n$ at constant speed would appear to be getting smaller and smaller and moving more and more slowly. She would never get there, of course; the boundary is "at infinity", not inside hyperbolic space.

Nonetheless, $\partial \mathbf{D}^n$ can be interpreted purely in terms of hyperbolic geometry as the visual sphere. For a given basepoint p in $\partial \mathbf{D}^n$, each hyperbolic ray from p, tends to a point on $\partial \mathbf{D}^n$. If q is another point in $\partial \mathbf{D}^n$, each line of sight from q appears, as seen from p, to trace out a segment of a great circle in the visual sphere of p, since p and the ray determine a hyperbolic two-plane. This visual segment converges to a point in the visual sphere of p; in this way, the visual sphere of q is mapped to the visual sphere at p. The endpoint of a line of sight from p, as seen by q, gives the inverse map. In this way the visual spheres of all observers in hyperbolic space can be identified, This construction is independent of the model, and so associates to hyperbolic space $\mathbf{H}^n$ the *sphere at infinity* $\mathbf{S}^{n-1}_\infty$.

The *stereographic projection* from an n-dimensional proper sphere $\mathbf{S} \subset \mathbf{E}^{n+1}$ onto a plane tangent to $\mathbf{S}$ at x is the map taking each point $p \in \mathbf{S}$ to the intersection q of the line $\overrightarrow{px'}$ with the plane, where x' is the point opposite x on $\mathbf{S}$.

Example 2.1. (Stereographic projection) The homeomorphism h: $\mathbf{E}^n \bigcup \infty \to \mathbf{S}^n$ can be chosen in such a way that it maps circles to circles and lines to circles minus $h(\infty)$. Stereographic projection can be extended to an inversion. Consequently, it is conformal, and takes spheres to spheres.

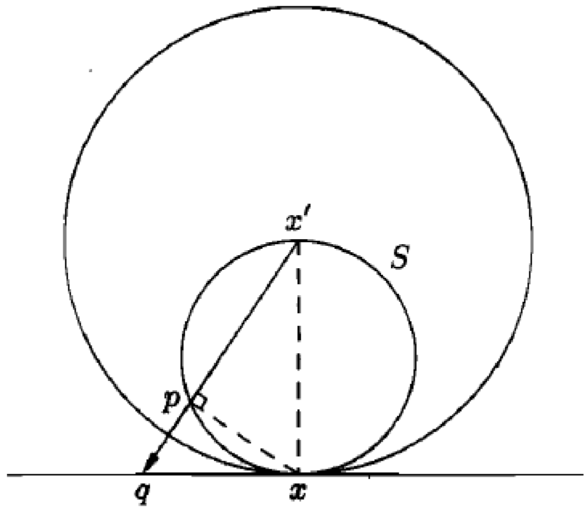

Figure 4: Stereographic projection from a sphere to a plane is identical to inversion in a sphere of twice the radius.

Our next model of $\mathbf{H}^n$ is derived from the Poincaré ball model by stereographic projection. We place the Poincaré ball $\mathbf{D}^n$ on the plane $\{x_0 = 0\}$ of $\mathbf{E}^{n+1}$, surrounded by the unit sphere $\mathbf{S} \subset \mathbf{E}^{n+1}$, and we project from $\mathbf{D}^n$ to the northern hemisphere of $\mathbf{S}^n$ with center at the south pole $(-1, 0, \ldots, 0)$, as shown in Figure 5 (a). This is an inverse stereographic projection, at least up to a dilatation (since the projection plane is equatorial rather than tangent).

In this way we transfer the geometry from the equational disk to the northern hemisphere to get the *hemisphere model*. Since stereographic projection is conformal and takes circles to circles, the hemisphere model is conformal and its geodesics are semicircles orthogonal to equator $\mathbf{S}^{n-1} = \partial \mathbf{D}^n$.

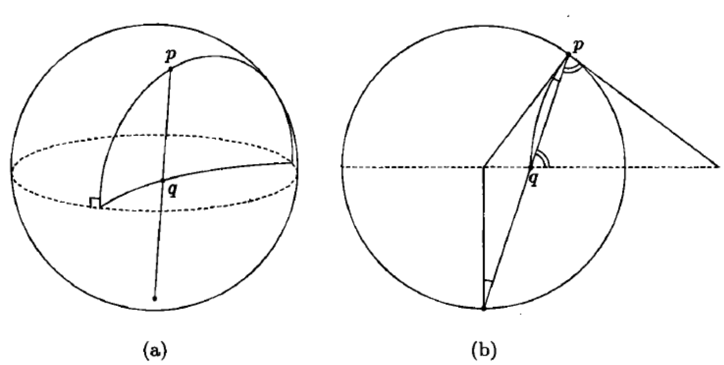

Figure 5: The hemisphere model. (a) By stereographic projection from the south pole of a sphere we map the equational disk to the northern hemisphere. Transfering the Poincaré disk metric by this mao we get a metric on the northern hemisphere whose geodesics are semicircles perpendicular to the equator. (b) The circle going through p and q and orthogonal to the equatorial disk is also orthogonal to the sphere. This shows that the projection of part (a) can also be obtained by following hyperbolic geodesics orthogonal to the equational disk.

It's easy to see from Figure 5 (b), that for each point $q \in \mathbf{D}^n$, the circle orthogonal to the equational disk $\mathbf{D}^n \subset \mathbf{D}^{n+1}$ and to $\mathbf{S}^n = \partial \mathbf{D}^{n+1}$ meets the northern hemisphere at the same point p as the image of q under the projection above. For each $q \in \mathbf{H}^n$, the hyperbolic ray from q perpendicular to $\mathbf{H}^n$ and pointing into the half-space we chose converges to a point in the corresponding visual hemisphere, so we get a map $\mathbf{H}^n \to \mathbf{S}^n_\infty$. By making $\mathbf{H}^n = \mathbf{D}^n$ be the equational disk in the

Poincaré ball model of $\mathbf{H}^{n+1} = \mathbf{D}^{n+1}$, we see that this map $\mathbf{H}^n \to \mathbf{S}^n_\infty$ coincides with the projection above.

From the hemisphere model we get the third important inversive model of hyperbolic space, also by stereographic projection.

Geodesics are given by semicircles orthogonal to the bounding plane $\mathbf{E}^{n-1}$ (Figure 6), and hyperbolic reflections are inversions in spheres orthogonal to the bounding plane. Clearly, this model, too, is conformal.

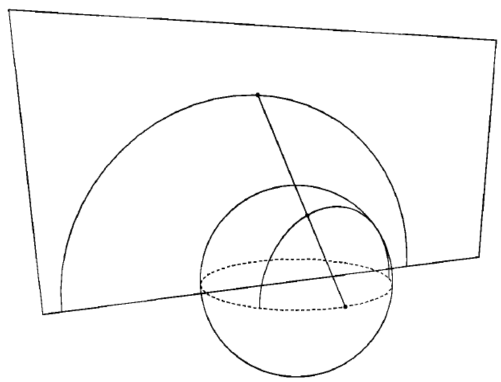

Figure 6: Geodesics in the upper half space model of hyperbolic space appear as semicircles orthogonal to the bounding plane, or half-lines perpendicular to it.

Figure 7 shows the same congruent tracts as Figure 3, but seen in the upper half-space model.

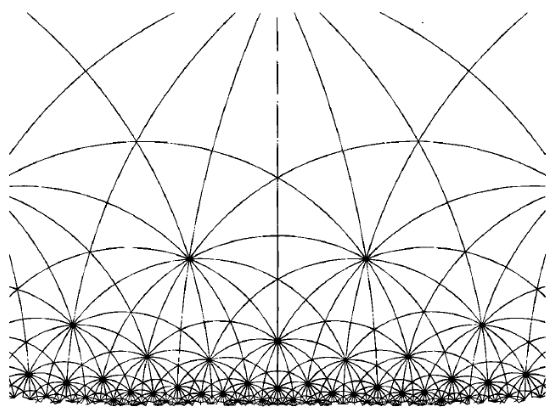

Figure 7: Another view of the hyperbolic world divided into congruent tracts. Upper half-plane projection

Example 2.2. (Euclidean similarities are hyperbolic isometries). A *similarity* of $\mathbf{E}^n$ is a transformation that multiplies all distances by the same (non-zero) factor. Any similarity of $\mathbf{E}^n$ can be composed from an element $O(n)$ (where we fix an origin for $\mathbf{E}^n$ aribitrarily), an expansion or contraction by a scalar factor, and a translation.

A similarity of $\mathbf{E}^{n-1}$ extends in a unique way to a similarity preserving upper half-space. By expressing it as a composition of reflections, shows that such a similarity is a hyperbolic isometry, .

The easy visibility of this significant subgroup of isometries of $\mathbf{H}^n$ is a frequently useful aspect of the upper half-space model.

Definition 2.2. A transformation of $\mathbf{S}^{n-1}_\infty$ that can be expressed as a composition of inversions is known as a Möbius transformation, and the group of all such transformations is the Möbius group, denote $M\ddot{o}b_{n-1}$. All hyperbolic isometries can be generated by reflections, it follows that the group of isometries of $\mathbf{H}^n$ is isomorphic to $M\ddot{o}b_{n-1}$.

Analyze and become familiar with the Möbius group.

(a) The subgroup of the Möbius group that fixes ∞ is isomorphic to the group of Euclidean similarities.

(b) The subgroup of the Möbius group $M\ddot{o}b_n$ that takes an $(n-1)$-sphere to iteself and fixes a point not on that sphere is isomorphic to the group $O(n)$.

(c) For $n > 1$, the Möbius group consists exactly of those homeomorphisms of $\mathbf{S}^n_\infty$ that take $(n-1)$-spheres to $(n-1)$-spheres.

(d) Any Möbius transformation that take a sphere $\mathbf{S}$ to a sphere $\mathbf{R}$ conjugates i_S to i_R.

(e) The group of the Möbius group that takes a k-sphere to itself is isomorphic to $M\ddot{o}b_k \times O(n-k)$.

(f) There is a subgroup of the Möbius group isomorphic to $O(n+1)$.

Property 2.2. (Minimal Hyperbolic Properties) In the discussion of hyperbolic geometry above, there was no attempt to characterize hyperbolic geometry using a minimal amount of structure. Here are some steps in this direction:

(a) Hyperbolic lines can be characterized in terms of the metric as curves that minimize distance between any two points. (Hint: in the upper half-space model, reduce to the case that p and q are on a verticle line.)

(b) We can characterize hyperbolic lines directly in terms of the group of isometries, as fixed-point sets.

(c) The only diffeomorphisms of upper half-space to itself that take all hyperbolic lines to hyperbolic lines are hyperbolic isometries. (This contrasts with the Euclidean case, where affine transformations take lines to lines.)

(d) The measure of angle also sufficient to define hyperbolic geometry.

3 The Hyperboloid Model and the Klein Model

A sphere in Euclidean space with radius r has constant curvature $1/r^2$. By analogy, since hyperbolic space has constant curvature -1, hyperbolic space should be a sphere of radius $i = \sqrt{-1}$.

To get an n-sphere, we start with the positive definite quadratic form Q^+ on $\mathbf{R}^{n+1}$ given by $Q^+(x) = x_0^2 + x_1^2 + \cdots + x_n^2$, where $x = (x_0, \ldots, x_n)$. This gives $\mathbf{R}^{n+1}$ a Euclidean metric $dx^2 = dx_0^2 + dx_1^2 + \cdots + dx_n^2$, making it into $\mathbf{E}^{n+1}$. Restricting to the unit sphere $\mathbf{S} = \{Q^+ = 1\}$, we get a Riemannian metric of constant positive curvature 1. The isometries of $\mathbf{S}$ come from linear transformations of $\mathbf{E}^{n+1}$ preserving Q^+; the group of these orthogonal transformations is denoted $O(n+1)$.

Let's start with the indefinite metric

$$ds^2 = -dx_0^2 + dx_1^2 + \cdots + dx_n^2 \tag{3.1}$$

in $\mathbf{R}^{n+1}$ associated to the quadratic form $Q^-(x) = -x_0^2 + x_1^2 + \cdots + x_n^2$. With this metric, $\mathbf{R}^{n+1}$ is often referred to as Lorentz space, and denoted $\mathbf{E}^{n,1}$. When $n = 3$, this is the universe of special relativity, although physicists usually reverse the sign of Q^-. In this interpretation, the vertical direction x_0 represents time, and the horizontal directions represent space. A vector is space-like, time-like or light-like depending on whether $Q^-(x)$ is positive, negative, or zero. Rays in *light cone* means $\{Q^- = 0, x_0 > 0\}$.

To identify antipodal points of $\mathbf{H}$, get a subset of projective space $\mathbf{RP}^n$. Unlike the case of the sphere, here antipodal points lie in disjoint components of $\mathbf{H}$, so this subset of $\mathbf{RP}^n$ can be modeled by one component of hyperboloid, say, the upper sheet $\mathbf{H}^+$, where $x_0 > 0$. This is the *hyperboloid model* of hyperbolic space.

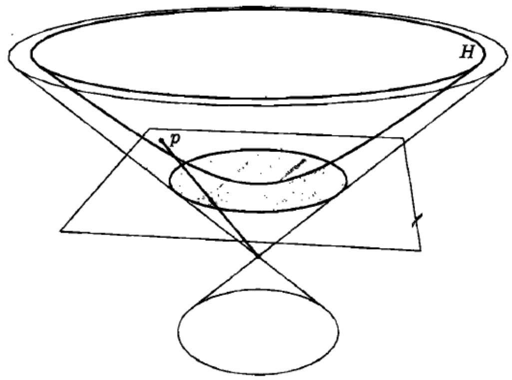

Figure 8: The hyperboloid model and the Klein model. A point $p = (x_0, x_1, \dots, x_n)$ on the hyperboloid maps to a point $(x_1/x_0, \dots, x_n/x_0)$ in $\mathbf{R}^n$, shown here as the horizontal hyperplane ($x_0 = 1$). This transfers the metric from the hyperboloid to the unit disk in $\mathbf{R}^n$, giving the projective model, or Klein model, for hyperbolic space.

In $\mathbf{E}^{n,1}$, the inner product is given by $-x_0y_0 + x_1y_1 + \cdots + x_ny_n$. The orthogonal complement of any non-zero vector x is an n-dimensional subspace, denoted by $x^\perp$; it contains x if and only if $Q^-(x) = 0$. The *orthogonal complement* of a subspace is the intersection of the orthogonal complements of its points.

Example 3.1. (Parametrization of geodesics) L is the intersection of the geodesic line through $p \in \mathbf{H}^+$ in direction v of $\mathbf{H}^+$. If v has unit length, L is parameterized with velocity 1 by $p \cosh t + v \sinh t$.

For the Klein model, L is a segment of straight line, meaning that this model is projectively correct: geodesics look straight. This makes the Klein model particularly useful for understanding incidence in a configuration of lines and planes. The sphere at infinity is just the unit sphere $\mathbf{S}^{n-1}_\infty$, the image in $\mathbf{RP}^n$ of the light cone. Angles are distorted in the Klein model, but they can be accurately and conveniently computed in the hyperboloid if one remembers to use the Lorentz metric of Equation 3.1, rather than the Euclidean metric.

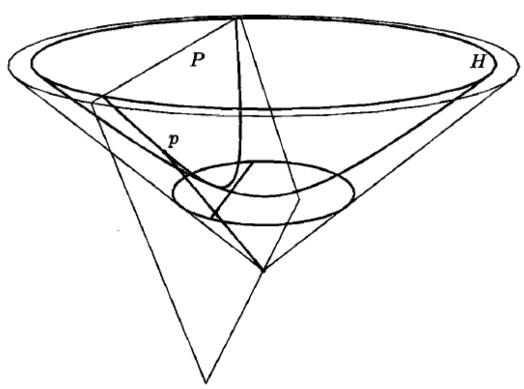

Figure 9: Geodesics in the hyperboloid and Klein models. In the hyperboloid model geodesics are the intersections of two-planes through the origin with the hyperboloid. In the projective model they're straight line segment.

Exhibit a correspondence between the Klein model and the hemisphere model that takes geodesics into geodesics.

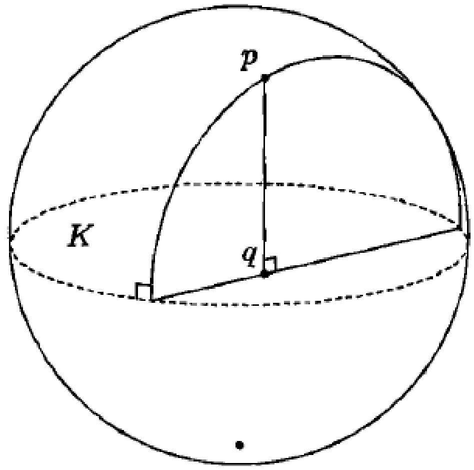

Figure 10: Going from the hemisphere to the Klein model. We get the Klein model of hyperbolic space from the hemisphere model by (Euclidean) orthogonal projection onto the equatorial disk. Compare Figure 5(a).

Consider points in $\mathbf{RP}^n$ that lie outside the closed ball. If $x \in \mathbf{RP}^n$ is such a point, Q^- is positive on the associated line $X \subset \mathbf{E}^{n,1}$. This means that Q^- is indefinite on the orthogonal complement $X^\perp$ of X, and that the corresponding hyperspace $x^\perp \subset \mathbf{RP}^n$ intersects hyperbolic space. We call $x^\perp$ the *dual hyperspace* of x. The hyperbolic significance of projective duality is that any line from x to $x^\perp$ is perpendicular to $x^\perp$. This is best seen in the hyperboloid model, as shown in the Figure 11(a): if $p \in X^\perp \bigcap \mathbf{H}^+$ represents a point in $x^\perp$ and $v \in X^\perp$ is any tangent vector at p that represents a direction in $x^\perp$, we want to show that v is perpendicular to the tangent vector w that represents the direction from p to X. But w lies in the plane determined by p and X, and is orthogonal to p; since p is orthogonal to X, this implies that w is in fact parallel to X, and consequently orthogonal to v.

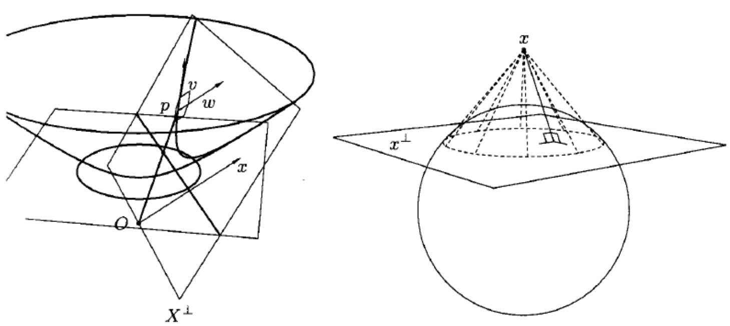

Figure 11: Duality between a hyperplane and a point. The dual of a point x outside $\mathbf{H}^n$ is a hyperplane $x^\perp$ intersecting $\mathbf{H}^n$. (a) Lines from $x^\perp$ to x are perpendicular to $x^\perp$, and lines perpendicular to $x^\perp$ go through x. (b) In $\mathbf{RP}^n$, the point x is the vertex of the cone tangent to $\mathbf{S}^{n-1}_\infty$ at $(n-2)$-sphere where $x^\perp$ meets $\mathbf{S}^{n-1}_\infty$

Definition 3.1. Consider $\mathbf{RP}^n$, if the intersection is inside $\mathbf{S}^{n-1}_\infty$, the lines meet in the conventional case, from the point of view of a hyperbolic observer. If the intersection is on $\mathbf{S}^{n-1}_\infty$, the lines converge together on the visual circle of the observer, and they are called *parallels*. Otherwise, they are called *ultraparallels*, and have a unique common perpendicular in $\mathbf{H}^n$, dual to their intersection point outside $\mathbf{S}^{n-1}_\infty$.

Definition 3.2. (Projective Transformations of Hyperbolic Space) A projective transformation is a self-map of $\mathbf{RP}^n$ obtained from an invertible linear map of $\mathbf{R}^{n+1}$ by passing to the quotient. An orthogonal transformation of $\mathbf{E}^{n+1}$, clearly gives rise to a projective transformation taking $\mathbf{S}^{n-1}_\infty$ to itself, the converse is also true.

This implies that any projective transformation of $\mathbf{RP}^n$ that leaves $\mathbf{H}^n$ invariant is an isometry, in contrast with the Euclidean situation, where there are many projective transformations that are not isometries: the affine transformations.

Example 3.2. (Consider shapes of Euclidean polygons.) The angles of a regular pentagon in plane Euclidean geometry are all 108°, but not all pentagons with 108° angles are regular. Consider the space of (not necessarily simple) pentagons having 108° angles and sides parallel to corresponding sides of a model regular pentagon, and sides parallel to the corresponding sides of a model regular pentagon, and parameterize this space by the (signed) side lengths $s_1, \cdots, s_5$.

4 Some Computations in Hyperboloid Space $\mathbf{H}^3$

We start with any triple (v_1, v_2, v_3) of unit vectors lying in $\mathbf{S}^2 \subset \mathbf{E}^3$. If they are linearly independent, so that no great circle, or spherical line, contains all three, they determine a spherical triangle, formed by joining each pair v_i, v_j by a spherical line segment of length $d(v_i, v_j) = \theta_{i,j} < \pi$. The dual basis to (v_1, v_2, v_3) is another triple (w_1, w_2, w_3) of vectors, but not necessary until vectors in $\mathbf{E}^3$, defined by the conditions $v_i \cdot w_i = 1$ and $v_i \cdot w_j = 0$ if $i \neq j$, for $i, j = 1, 2, 3$. If we let V and W be the matrices with columns v_i and w_i, this can be expressed as $W^t V = I$.

Geometrically, w_i points in the direction of the normal vector of the plane spanned by v_j and v_k, where i, j and k are distinct; it follows that the angle Φ_i of the spherical triangle $v_1 v_2 v_3$ at v_i is $\pi - \angle(w_j, w_k)$; since the angle between two planes is the supplement of the angle between their outward normal vectors.

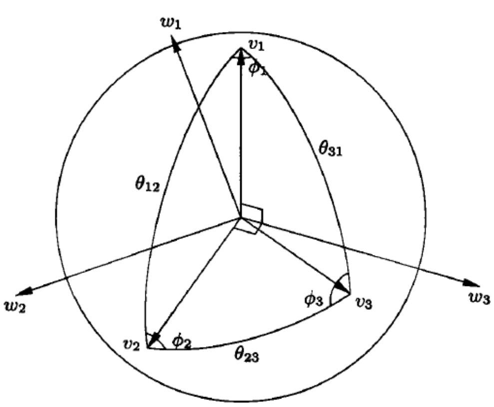

Figure 12: Proving the spherical law of cosines

To relate all these angles, we consider the matrices $V^t V$ and $W^t W$ of inner products of the two bases, and notice that they are inverse to one another and that

$$V^t V = \begin{pmatrix} 1 & c_{12} & c_{13} \\ c_{12} & 1 & c_{23} \\ c_{13} & c_{23} & 1 \end{pmatrix} \tag{4.1}$$

where $c_{ij} = v_i \cdot v_j = \cos\theta_{ij}$. Thus $W^t W = (V4V)^{-1}$ is a multiple of the matrix of cofactors of $V^t V$

$$W^t W = \frac{1}{det(V^t V)} = \begin{pmatrix} 1 - c_{23}^2 & c_{13}c_{23} - c_{12} & c_{12}c_{23} - c_{13} \\ c_{13}c_{23} - c_{12} & 1 - c_{13}^2 & c_{12}c_{13} - c_{23} \\ c_{12}c_{23} - c_{13} & c_{12}c_{13} - c_{23} & 1 - c_{12}^2 \end{pmatrix} \tag{4.2}$$

From this we can easily compute, say,

$$\cos\phi_3 = -cos\angle(w_1, w_2) = -\frac{w_1 \cdot w_2}{|w_1||w_2|} = \frac{\cos\theta_{12} - \cos\theta_{13}\cos\theta_{23}}{\sin\theta_{13}\sin\theta_{23}} \tag{4.3}$$

or, on the more familiar notation where A, B, C stands for the angles at v_1, v_2, v_3 and a, b, c stand for the opposite sides,

$$\cos c = \cos a \cos b + \sin a \sin b \cos C. \tag{4.4}$$

This is called the spherical law of cosines. The dual spherical law of cosines is obtained by reserving the roles of (v_1, v_2, v_3) and (w_1, w_2, w_3) with respect to the triangle:

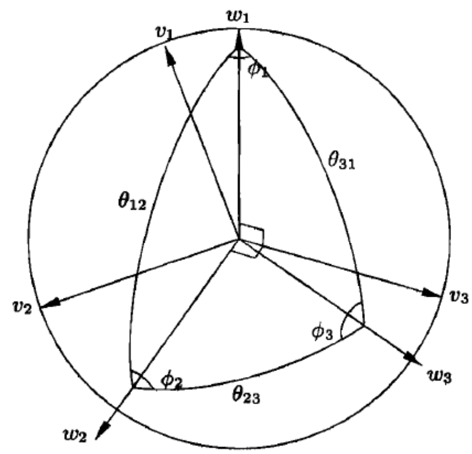

Figure 13: The dual spherical law of cosines

we set the v_i not the vertices of the triangle, but to unit vectors orthogonal to the planes containing the sides. Then $\phi_k = \pi - \angle(v_i, v_j)$ for i, j, k distinct, and $\theta_{ij} = \angle(w_i, w_j)$. We obtain, for example,

$$\cos \theta_{12} = \cos \angle(w_1, w_2) = \frac{w_1 \cdot w_2}{|w_1||w_2|} = \frac{\cos \phi_2 \cos \phi_1 + \cos \phi_3}{\sin \phi_2 \sin \phi_1} \tag{4.5}$$

or

$$\cos C = -\cos A \cos B + \sin A \sin B \cos c. \tag{4.6}$$

Turn to Lorentz space $\mathbf{E}^{2,1}$, the non-zero vectors can have real, zero or imaginary length. To cut down the number of cases, we consider only vectors of non-zero length. We may assume that they are normalized in the sense that they have length 1 or i and their x_0-coordinate is positive if they have length i. We recall from last section that if a normalized vector $x \in \mathbf{E}^{2,1}$ has length i, it stands for a point on the hyperboloid model $\mathbf{H}^+$ of the hyperboloid plane, just as a unit vector in $\mathbf{E}^3$ gives a point in $\mathbf{S}^2$. If x has length 1, it lies outside the hyperbolic plane, and we denote by $x^\perp$ the trace in the hyperbolic plane of its dual line.

Suppose x and y are normalized and distinct. The quadratic form Q^-, restricted to the plane spanned by x and y, can have signature $(2, 0)$, $(1, 0)$ or $(1, 1)$, corresponding to the cases where the plane intersects the hyperboloid, is tangent to the cone at infinity, or avoids both.

Proposition 4.1. (interpretation of the inner product) If x and y are normalized vectors of non-zero length in $\mathbf{E}^{n,1}$, either

(a) $x, y \in \mathbf{H}^+$ have length i, and $x \cdot y = -\cosh d(x, y)$; or

(b) $x \in \mathbf{H}^+$ has length i, y has length 1 and $x \cdot y = \pm \sinh d(x, y^\perp)$;

(c) x and y have length 1, and the hyperplanes $x^\perp, y^\perp \subset \mathbf{H}^n$ are secant, parallel or ultraparallel depending on whether Q^- has signature $(2, 0)$, $(1, 0)$ or $(1, 1)$ on the plane spanned by x and y. In the first case, $x \cdot y = \pm \cos \angle(x^\perp, y^\perp)$; in the second, $x \cdot y = \pm 1$; and in the third, $x \cdot y = \pm \cosh d(x^\perp, y^\perp)$.

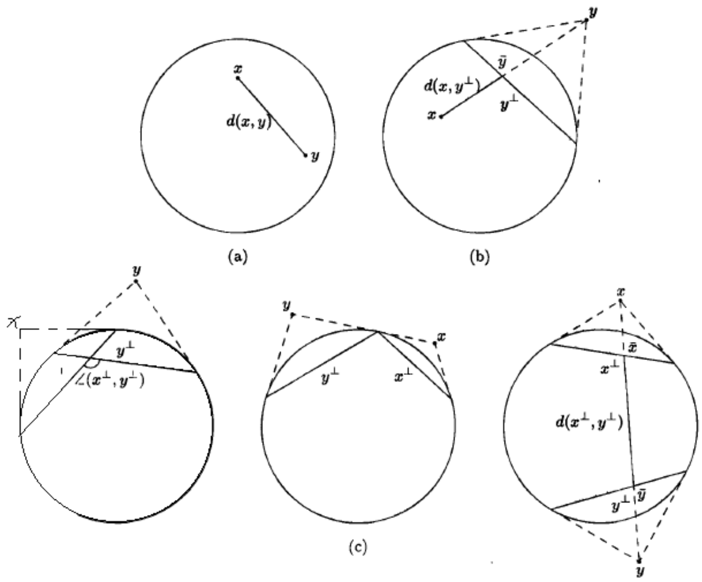

Figure 14: Interpretation of the inner product for various relative positions of two points. The labels correspond to the cases in Proposition 4.1; the figures are drawn in the projective model.

Proof. Let $\mathbf{P}$ be the plane spanned by x and y. In cases (a) and (b), $\mathbf{P}$ intersects $\mathbf{H}^+$ in a hyperbolic line $L \ni x$, and by Example 3.1 this line is parameterized with velocity 1 by $x \cosh t + v \sinh t$, where v is a unit tangent vector to $\mathbf{H}^+$ at x.

If $y \in \mathbf{H}^+$, this implies that $y = x \cosh t + v \sinh t$ for $t = \pm d(x, y)$, depending on the way we chose v. Since x and v are orthogonal, we get

$$x \cdot y = x \cdot (x \cosh t + v \sinh t) = -\cosh t = -\cosh d(x, y)$$

If, on the other hand, $y \notin \mathbf{H}^+$, Example 4.1 shows that the distance $d(x, y^\perp)$ is achieved for the point $\overline{y} = L \bigcap y^\perp$, because L is the unique perpendicular from x to $y^\perp$. Thus $\overline{y} = x \cosh t + v \sinh t$ for $t = d(x, y^\perp)$, and y, being a linear combination of x and v orthogonal to $\overline{y}$, must be of the form $\pm(x \sinh t + v \cosh t)$ (recall that x and y are normalized.) We conclude that

$$x \cdot y = \pm x \cdot (x \sinh t + v \cosh t) = \pm \sinh t = \pm \sinh d(x, y^\perp).$$

The third possibility in (c) is variation on (a) and (b). Here $L = \mathbf{P} \bigcap \mathbf{H}^+$ contains neither x nor y, but we can parameterize it starting at $\overline{x} = L \bigcap x^\perp$. Then $x = \pm v$, $\overline{y} = L \bigcap y^\perp = x^\perp \cosh t + v \sinh t$ for $t = d(x^\perp, y^\perp)$, and $y = \pm(\overline{x} \sinh t + v \cosh t)$, so $x \cdot y = \pm \cosh d(x^\perp, y^\perp)$.

We are left with the first two possibilities in (c). If Q^- is positive definite on $\mathbf{P}$, it is indefinite on the orthogonal complement $\mathbf{P}^\perp$, so $\mathbf{P}^\perp \bigcap \mathbf{H}^+ = x^\perp \bigcap y^\perp$ is non-empty. Let p be a point in this intersection; to measure $\cos \angle(x^\perp)$ it is enough to find tangent vectors to $\mathbf{H}^+$ at p that are normal to $x^\perp$ and $y^\perp$, and take the cosine of their angle. But x and y themselves conserve as much as such tangent vectors, so $\cos \angle(x^\perp, y^\perp) = \pm x \cdot y$.

If Q^- is positive semidefinite on $\mathbf{P}$, it is also positive semidefinite on $\mathbf{P}^\perp$, so $\mathbf{P}^\perp \bigcap \mathbf{H}^+ = x^\perp \bigcap y^\perp$ is empty, but $\mathbf{P}^\perp \bigcap \mathbf{S}^{n-1}_\infty$ consists of a single line through the origin. Thus $x^\perp$ and $y^\perp$ meet at infinity–they are parallel. The value of $x \cdot y$ follows from the fact that this case is a limit between the previous two. $\square$

Example 4.1. (minimum distance implies perpendicularity)

(a) If $L \subset \mathbf{H}^n$ is a line, $y \in \mathbf{H}^n$ is a point outside L and x is a point on L such that the distance $d(x, y)$ is minimal, the line xy is perpendicular to L.

(b) If $L, M \subset \mathbf{H}^n$ are non-intersecting lines and $x \in L$ and $y \in M$ are points on L and M such that the distance $d(x, y)$ is minimal, xy is perpendicular to L and M.

Now we calculate the trigonometric formulas for a triangle in $\mathbf{H}^2$, or, more generally, the intersection with $\mathbf{H}^2$ of a triangle in $\mathbf{RP}^2$. As before, we let (v_1, v_2, v_3) be a basis of normalized vectors in $\mathbf{E}^{2,1}$, forming a matrix V, and we look at its dual basis (w_1, w_2, w_3), whose vectors from a matrix W. Here, V and W are no longer inverse to each other; instead, we can write $W^tSV = I$, where S is a symmetric matrix expressing the inner product associated with Q^- in the canonical basis-here the diagonal matrix with diagonal entries $(-1, 1, 1)$. However, the matrices of inner products, V^tSV and W^tSW, are still inverse to each other:

$$(V^tSV)(W^tSW) = (V^tsV)(V^{-1}W) = V^tSW = (W^tSV) = I$$

Since some of the v_i may have imaginary length, V^tSV no longer has all ones in the diagonal; instead, it looks like this:

$$V^tSV = \begin{pmatrix} \varepsilon_1 & c_{12} & c_{13} \\ c_{12} & \varepsilon_2 & c_{23} \\ c_{13} & c_{23} & \varepsilon_3 \end{pmatrix} \tag{4·7}$$

where $\varepsilon_i = v_i \cdot v_i = \pm 1$. It follows, as before, that matrix of inner products of w_i is

$$W^tSW = \frac{1}{\det V^tSV} = \begin{pmatrix} \varepsilon_2\varepsilon_3 - c_{23}^2 & c_{13}c_{23} - \varepsilon_3 c_{12} & c_{12}c_{23} - \varepsilon_2 c_{13} \\ c_{13}c_{23} - \varepsilon_3 c_{12} & \varepsilon_1\varepsilon_3 - c_{13}^2 & c_{12}c_{13} - \varepsilon_1 c_{23} \\ c_{12}c_{23} - \varepsilon_2 c_{13} & c_{12}c_{13} - \varepsilon_1 c_{23} & \varepsilon_1\varepsilon_2 - c_{12}^2 \end{pmatrix} \tag{4·8}$$

Proposition 4.2. (Ideal Triangles) All ideal triangles are congruent, and have area π.

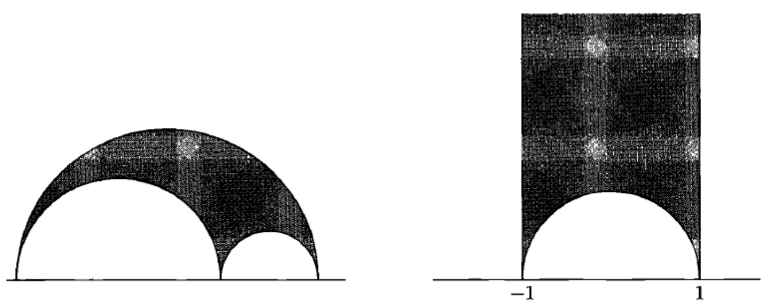

Figure 15: All ideal triangles are congruent. Given any ideal triangle we can send one of its vertices to ∞ by inversion, then apply a Euclidean similarity to send the remaining two vertices to $(-1, 0)$ and $(1, 0)$.

Proof. Using the upper half-plane model of $\mathbf{H}^2$, it is easy to see that any ideal triangle can be transformed by isometries so as to match a model triangle with vertices ∞, $(-1, 0)$, $(1, 0)$.

Now let the coordinates of the upper half-plane be x and y, with the x-axis as the boundary. The model triangle is the region given by $-1 \leq x \leq 1$ and $y \geq \sqrt{1 - x^2}$, with hyperbolic area element $(1/y^2)dxdy$. Thus the area is

$$\int_{-1}^{1} \int_{\sqrt{1-x^2}}^{\infty} \frac{1}{y^2} dy dx = \int_{-1}^{1} \frac{1}{\sqrt{1 - x^2}} dx = \int_{-\pi/2}^{\pi/2} \frac{1}{\cos\theta} d\theta = \pi$$

□

Proposition 4.3. (Area of Hyperbolic Triangles) The area of a hyperbolic triangle is π minus the sum of the interior angles (the angle being zero for a vertex at infinity).

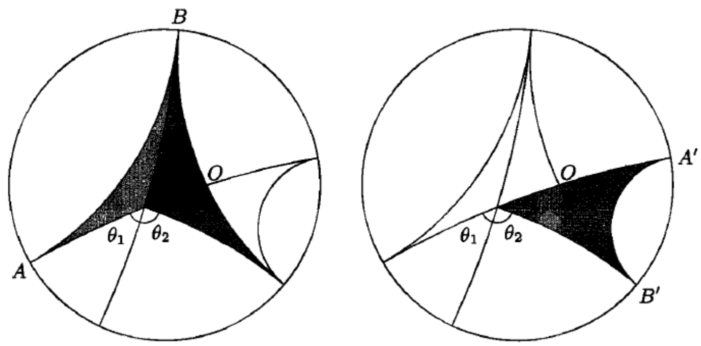

Figure 16: Area of $\frac{2}{3}$-ideal triangles. By definition, the areas of the shaded triangles on the left are $A(\theta_1)$ and $A(\theta_2)$. Likewise, the area of the shaded triangle on the right is $A(\theta_1 + \theta_2)$. But the shaded areas in the two figures coincide, because the triangles OAB and $OA'B'$ are congruent by a reflection through O. Therefore $A(\theta)$ is an additive function of θ; this is used to compute the area of $\frac{2}{3}$-ideal triangles.

Proof. When all angles are zero we have an ideal triangle. We next look at $\frac{2}{3}$-ideal triangles, those with two vertices at infinity. Let $A(\theta)$ denote the area of such a triangle with angle $\pi - \theta$ at the finite vertex. This is well-defined because all $\frac{2}{3}$-ideal triangles with the same angle at the finite vertex are congruent-the reasoning is similar to that for ideal triangles.

Gauss's key observation is that A is an additive function, that is, $A(\theta_1 + \theta_2) = A(\theta_1) + A(\theta_2)$, for $\theta_1, \theta_2, \theta_1 + \theta_2 \in (0, \pi)$. The proof of this follows from Figure 16. It follows that A is a Q-linear function from $(0, \pi)$ to $\mathbf{R}$. It is also continuous, so it must be $\mathbf{R}$-linear. But $A(\pi)$ is the area of an ideal triangle, which is π by Proposition 4.2; it follows that $A(\theta) = \theta$, and the area of $\frac{2}{3}$-ideal triangle is the complement of the angle at the finite vertex.

A triangle with two or three finite vertices can be expressed as the difference between an ideal triangle and two or three $\frac{2}{3}$-ideal ones, as shown in Figure 17. You can check the details. $\square$

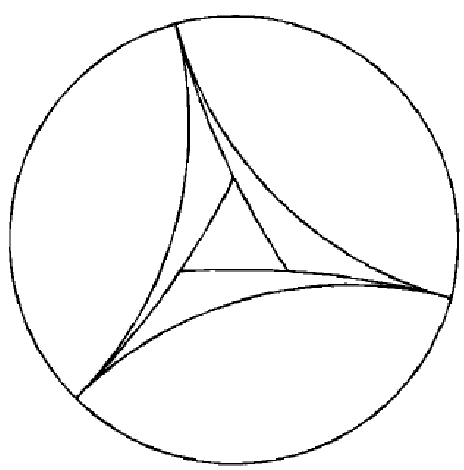

Figure 17: Area of general hyperbolic triangles. If you subtract a finite hyperbolic triangle from a suitable ideal triangle, you get three $\frac{2}{3}$-triangles. Adding up angles and areas gives Proposition 4.3

5 Hyperbolic Isometries

Let $g : \mathbf{H}^3 \to \mathbf{H}^3$ be an orientation-preserving isometry other than the identity. An axis of g is any line L that is invariant under g and on which g acts as a (possibly trivial) translation.

Proposition 5.1. (Axis is unique) A non-trivial orientation-preserving isometry of $\mathbf{H}^3$ can have at most one axis.

Proof. Suppose that L and M are distinct axes for an orientation-preserving isometry g. If g fixes both L and M pointwise, take a point x on M but not on L. Then g fixes the plane containing L and x, because it fixes three non-collinear pints on it. Since g preserves orientation, it is the identity.

If, on the other hand, L is translated by g, we have $d(x, M) = d(g(x), M)$ for $x \in L$, so the function $d(x, M)$ is periodic and therefore bounded. But two distinct lines cannot remain a bounded distance from each other in both directions, since that would imply thy have the same two endpoints on the sphere at infinity. $\square$

Any orientation-preserving isometry of $\mathbf{E}^3$ is either a translation, a rotation about some axis, or a *screw motion*, that is, a rotation followed by a translation along the axis of rotation. The situation in $\mathbf{H}^3$ is somewhat richer, and has its own special terminology.

If a non-trivial orientation-preserving isometry g of $\mathbf{H}^3$ has an axis that is fixed pointwise, it is called an *elliptic isometry*, or a *rotation* about its axis. In this case the orbit of a point p off the axis-the set of points $g^k(p)$, for $k \in \mathbf{Z}$-lies on a circle around the axis.

If g has an axis that is translated by non-trivial amount, it is called *hyperbolic*. There are two possibilities here: the orbit of a point off the axis may lie on a plane, always on the same side of the axis; it is in fact contained in an equidistant curve. In this case we say that g is a *translation*. Alternatively, the orbit can be the vertices of a polygonal helix centered around the axis; in this case g is a *screw motion*, as can be seen by applying a compensatory translation.

By the proposition above, no transformation can be at the same time elliptic and hyperbolic. But there are isometries that are neither elliptic nor hyperbolic: they are called *parabolic*. For instance, any isometry of $\mathbf{H}^3$ that appears as a Euclidean translation parallel to the bounding plane in the upper half-space model is parabolic.

Lemma 5.1. (Common perpendicular for lines in $\mathbf{H}^3$) Two distinct lines in $\mathbf{H}^3$ are either parallel, or they have a unique common perpendicular.

Proof. Let the lines be X and Y, and consider the distance function $d(x, y)$ between points $x \in X$ and $y \in Y$. If the lines are not parallel, this function goes to ∞ as either x or y or both go to ∞; therefore it has a minimum, attained at points x_0 and y_0.

If the minimum is zero, that is, if the lines cross, any common perpendicular must go through intersection point, otherwise we'd have a triangle with two right angles, which is impossible by Proposition 4.3. As there is a unique line orthogonal to the plane spanned by X and Y and passing through their intersection point, the lemma is proved in this case. (Notice that this part is false in dimension greater than three.)

If the minimum distance is not zero, the line between x_0 and y_0 is a perpendicular by Example 4.1. If there were another common perpendicular, we'd obtain a quadrilateral in space with all right angles (although conceivably its sides could cross). Subdividing the quadrilateral by a diagonal, we'd get two plane triangles, whose angles add to at least 2π; this is again impossible. $\square$

If L is a line in $\mathbf{H}^3$, we denote by r_L the reflection in L, which is the rotation of π about L.

Proposition 5.2. (Finding the axis) Any non-trivial orientation-preserving isometry g of $\mathbf{H}^3$ can be written in the form $g = r_L \circ r_M$, where the lines L and M are parallel, secant or neither depending on whether g is parabolic, elliptic or hyperbolic. The axis of g is the common perpendicular of L and M, if it exists.

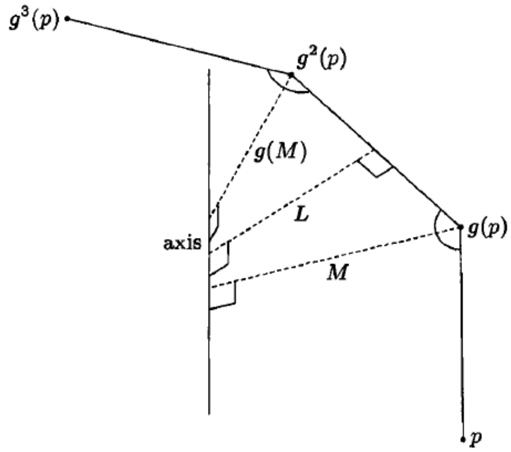

Figure 18: The axis of a three-dimensional isometry. The axis of an isometry g of $\mathbf{H}^3$ may be constructed, in the generic case, by connecting the orbit of a point p in a polygonal path.

Proof. Take any point p such that $g(p) \neq p$. If $g^2(p) = p$, the midpoint q of the line segment $\overline{pg(p)}$ is fixed by g, and the plane through q perpendicular to the line $pg(p)$ is invariant. Since g reverses the orientation of this plane, it must act on it as a reflection, fixing a line K. Therefore, $g = r_K$ is elliptic of order two. In this case, one can take L and M to be orthogonal lines, both orthogonal to K.

If $p, g(p)$ and $g^2(p) \neq p$ are collinear, p is fortuitously on the axis f g, and g is hyperbolic. We can replace p by some other point not on this axis, and reduce to the next case.

In the remaining case, we let M be the bisector of the angle $pg(p)g^2(p)$, so that r_M fixes $g(p)$ and interchanges p with $g^2(p)$. To define L, we look at the dihedron with the edge $g(p)g^2(p)$ whose sides contain p and $g^3(p)$, respectively, and take L as the line that bisects this dihedron and is also a perpendicular bisector of the segment $\overline{pg(p)}$, as shown in Figure 18.(The dihedral angle along $pg(p)$ may be 0 or π, but this doesn't cause problems.) By symmetry, r_L interchanges $g(p)$ with $g^2(p)$, and also p with $g^3(p)$. Therefore, $r_L \circ r_M$ sends p to $g(p)$, $g(p)$ to $g^2(p)$. Since $r_L \circ r_M$ and g agree at three non-collinear points, they agree on the whole plane containing these three points. Therefore they agree everywhere, since they both preserve orientation.

The common perpendicular of L and M, if it exists, is the axis of g, because it is invariant under $r_L \circ r_M$, which acts on it as a translation. The sorting into cases now follows from Lemma 5.1 and from the definitions of parabolic, elliptic and hyperbolic transformations. □

A *convex function* on a Riemannian manifold is a real-valued function f such that, for every geodesic γ, parameterized at a constant speed, the induced function $f \circ \gamma$ is convex. In other words, for every $t \in (0, 1)$,

$$f \circ \gamma(t) \leq tf \circ \gamma(0) + (1 - t)f \circ \gamma(1).$$

If the inequality is strict for all non-constant γ, we say that f is *strictly convex.*

Theorem 5.1. (Distance function is convex) The distance function $d(x, y)$, considered as a map $d : \mathbf{H}^n \times \mathbf{H}^n \to \mathbf{R}$, is convex. The composition $d \circ \gamma$ is strictly convex for any geodesic γ in $\mathbf{H}^n \times \mathbf{H}^n$ whose projections to the two factors are distinct lines.

Proof. William P. Thurston, Three-Dimensional Geometry and Topology, Princeton University Press, 1997, P91-93. □

The *translation distance* of an isometry $g : \mathbf{H}^n \to \mathbf{H}^n$ is the function $d_g(x) = d(x, g(x))$. By applying Theorem 5.1 to the graph of g, which is a geodesic-preserving embedding of $\mathbf{H}^n$ in $\mathbf{H}^n \to \mathbf{H}^n$, we get:

Corollary 5.1. (translation distance is convex) For any isometry g of $\mathbf{H}^n$, the translation distance d_g is a convex function on $\mathbf{H}^n$. It is strictly convex except along lines that map to themselves.

Property 5.1. (classification of isometries of $\mathbf{H}^n$) Let g be an isometry of $\mathbf{H}^n$.

(a) g is hyperbolic if and only if the infimum of d_g is positive. This infimum is attained along a line, which is the unique axis for g. The endpoints of the axis are fixed points of g of $\mathbf{S}^{n-1}_\infty$.

(b) g is parabolic if and only if the infimum of d_g is not attained. This infimum is then zero. g fixes a unique point p on $\mathbf{S}^{n-1}_\infty$, and acts as a Euclidean isometry in the upper half-space model with p at ∞.

(c) g is elliptic if and only if d_g takes the value zero. The set $d_g^{-1}(0)$ is a hyperbolic subspace of dimension k, for $0 \le k \le n$.

Proof. If d_g attains a positive infimum at some point x, it also attains the infimum at $g(x)$. By convexity, d_g has the same value on the line segment joining x and $g(x)$, so by Corollary 5.1 the line through x and $g(x)$ is invariant. This line is translated along itself, so g is hyperbolic. The uniqueness of its axis follows just as in the second half of the proof of Proposition 5.1. To show that the axis are the only fixed points on $\mathbf{S}^{n-1}_\infty$, we assume that one endpoint is at infinity in the upper half-space model. Then g acts as a Euclidean similarity on the bounding hyperplane $\mathbf{S}^{n-1}_\infty\ \{\infty\}$. As such it can have at most one fixed point, unless it is the identity.

If d_g does not attain an infimum, there is a sequence $\{x_i\}$ such that $d_g(x_i)$ tends toward the infimum. By compactness, we can assume that $\{x_i\}$ converges to a point $x \in \mathbf{S}^{n-1}_\infty$, which must be fixed by g. We can take $x = \infty$ in the upper half-space projection, so g acts as a Euclidean similarity. If this similarity has no fixed point on $\mathbf{S}^{n-1}_\infty\ \{\infty\}$, it is an isometry; therefore d_g goes to zero on any vertical ray, and $\inf d_g = 0$. Also, since g has no axis and no fixed point in $\mathbf{H}^n$, it is parabolic.

If, instead, g does fix a point on $\mathbf{S}^{n-1}_\infty\ \{\infty\}$, it leaves invariant the vertical line L through that point. If P is a (hyperbolic) hyperbolic orthogonal to L, the closed region F between P and $g(P)$ is a fundamental domain for g, that is, for any point $x \in \mathbf{H}^n$, there is some $k \in \mathbf{Z}$ such that $g^k(x) \in F$. In particular, any value of d_g is achieved inside F. Because d_g does not attain its infimum, the compactness argument of the preceding paragraph shows that g fixes a point in $\overline{F} \bigcap \mathbf{S}^{n-1}_\infty$. But if g fixes three points on $\mathbf{S}^{n-1}_\infty$, it fixes a whole plane in $\mathbf{H}^n$, contradicting the assumption that $\inf d_g$ is not attained.

Finally, if d_g takes the value of zero, g is by definition elliptic, and its zero-set is a k-dimensional subspace because the entire line joining any two fixed points is fixed. □

6 Complex Coordinates for Hyperbolic Three-Space $\mathbf{H}^3$

The complex plane $\mathbf{C}$ embeds naturally in the complex projective line $\mathbf{CP}^1$, the set of complex lines (one-dimensional complex subspaces) of $\mathbf{C}^2$. The embedding maps a point $z \in \mathbf{C}$ to the complex line spanned by $(z, 1)$, seen as a point inn $\mathbf{CP}^1$; we call z the *inhomogeneous coordinate* for this point, while any pair $(tz, t) \in \mathbf{C}^2$, with $t \in \mathbf{C}^* = \mathbf{C}\ \{0\}$, is called a set of *homogeneous coordinates* for it. The remaining point in $\mathbf{CP}^1$, namely the subspace spanned by $(1, 0)$, is the *point of infinity*; we can make ∞ its "inhomogeneous coordinate".

Topologically, $\mathbf{CP}^1$ is the one-point compactification $\hat{\mathbf{C}}$ of $\mathbf{C}$, so we can extend the usual i-dentification of $\mathbf{E}^2$ with $\mathbf{C}$ to ∞. This shows that $\mathbf{CP}^1$ is a topological sphere, called *Riemann sphere.*

As in the real case, a *projective transformation* of $\mathbf{CP}^1$ is what you get from an invertible linear map of $\mathbf{C}^2$ by passing to the quotient. Projective transformations are homeomorphisms of $\mathbf{CP}^1$. If a projective transformation A comes from a linear map with matrix

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \tag{6.1}$$

its expression in inhomogeneous coordinates is

$$A(z) = \frac{az + b}{cz + d} \tag{6.2}$$

(naturally, this should be interpreted as giving a/c for $z = \infty$ and ∞ for $z = -d/c$). A map $A : \mathbf{CP}^1 \to \mathbf{CP}^1$ of the form (6.2) (with $ad - bc \neq 0$) is called a *linear fractional transformation* (or *fractional linear transformation*). Linear fractional transformations are Möbius transformations and behave in a familiar way as follows:

(a) Under the usual identification $\mathbf{CP}^1 = \mathbf{S}^2$, any linear fractional transformation is a Möbius transformation, that is, a composition of inversions.

(b) Any orientation-preserving Möbius transformation of $\mathbf{S}^2$ is a linear fractional transformation.

Two non-singular linear maps of $\mathbf{C}^2$ give the same projective transformation of $\mathbf{CP}^1$ if and only if one is a scalar multiple of the other. Thus, identifying together scalar multiples in the linear group $\mathbf{GL}(2, \mathbf{C})$ gives the group of projective transformations of $\mathbf{CP}^1$, which we naturally denote by $\mathbf{PGL}(2, \mathbf{C}) = \mathbf{GL}(2, \mathbf{C})/\mathbf{C}^*$. This group is also known as $\mathbf{PSL}(2, \mathbf{C})$, because it can be obtained by identifying together scalar multiples in the special linear group $\mathbf{SL}(2, \mathbf{C})$, consisting of linear transformations of $\mathbf{C}^2$ with unit determinant.

A Möbius transformation of $\mathbf{S}^{n-1}_\infty$ can be extended to a unique isometry of $\mathbf{H}^n$. Since $\mathbf{PGL}(2, \mathbf{C})$ acts on $\mathbf{S}^2_\infty$ by Möbius transformations, this action can be extended to all of $\mathbf{H}^3$, providing the first link between hyperbolic geometry and the complex numbers:

Theorem 6.1. The group of orientation-preserving isometries of $\mathbf{H}^3$ is $\mathbf{PGL}(2, \mathbf{C})$, identified via the action on $\mathbf{S}^2_\infty = \mathbf{CP}^1$.

Reference

William P. Thurston, Three-dimensional geometry and topology, Princeton University Press, 1997