HYPERBOLIC GEOMETRY

JOUNI PARKKONEN

0. Warmup: Euclidean and spherical geometry

In this short section we will (re)acquaint ourselves with some of the basic concepts and objects of the rest of the course with the aid of Euclidean and spherical geometry which I assume are to some extent familiar to everyone.

0.1. **Metric spaces.** Recall that a function $d\colon X \times X \to [0, +\infty[$ is a *metric* in the nonempty set X if it satisfies the following properties

(1) $d(x, x) = 0$ for all $x \in X$ and $d(x, y) > 0$ if $x \neq y$,
(2) $d(x, y) = d(y, x)$ for all $x, y \in X$, and
(3) $d(x, y) \leq d(x, z) + d(z, y)$ for all $x, y, z \in X$ (the *triangle inequality*).

The pair (X, d) is a *metric space*. Open and closed balls in a metric space, continuity of maps between metric spaces and other "metric properties" are defined in the same way as in Euclidean space, using the metrics of X and Y instead of the Euclidean metric.

If (X_1, d_1) and (X_2, d_2) are metric spaces, then a map $i\colon X \to Y$ is an *isometric embedding*, if

$$d_2(i(x), i(y)) = d_1(x, y)$$

for all $x, y \in X_1$. If the isometric embedding i is a bijection, then it is called an *isometry* between X and Y.

The isometries of a metric space X form a group $\mathrm{Isom}(X)$, the *isometry group* of X, with the composition of mappings as the group law.

A map $i\colon X \to Y$ is a *locally isometric embedding* if each point $x \in X$ has a neighbourhood U such that the restriction of i to U is an isometric embedding. A (locally) isometric embedding $i\colon I \to X$ is

(1) a *(locally) geodesic segment*, if $I \subset \mathbb{R}$ is a (closed) bounded interval,
(2) a *(locally) geodesic ray*, if $I = [0, +\infty[$, and
(3) a *(locally) geodesic line*, if $I = \mathbb{R}$.

Note that in Riemannian geometry, the definition of a geodesic line is different from the above: in general a Riemannian geodesic line is only a locally geodesic line according to our definition.

0.2. **Euclidean space.** Let us denote the *Euclidean inner product* of $\mathbb{R}^n$ by

$$(x|y) = \sum_{i=1}^{n} x_i y_i \,.$$

The *Euclidean norm* $\|x\| = \sqrt{(x|x)}$ defines the *Euclidean metric* $d(x, y) = \|x - y\|$. The triple $\mathbb{E}^n = (\mathbb{R}^n, (\cdot|\cdot), \|\cdot\|)$ is n-dimensional *Euclidean space.*

Euclidean space is a *geodesic metric space*: For any two distinct points $x, y \in \mathbb{E}^n$, the map $j_{x,y}\colon \mathbb{R} \to \mathbb{E}^n$,

$$j_{x,y}(t) = x + t\frac{y - x}{\|y - x\|},$$

Last update: December 11, 2013.

is a geodesic line that passes through the points x and y. The restriction $j_{x,y}|_{[0,\|x-y\|]}$ is a geodesic segment that *connects* x to y: $j(0) = x$ and $j(\|x-y\|) = y$. In fact, this is the only geodesic segment that connects x to y up to replacing the interval of definition $[0, \|x-y\|]$ of the geodesic by $[a, a+\|x-y\|]$ for some $a \in \mathbb{R}$. More precisely: A metric space (X, d) is *uniquely geodesic*, if for any $x, y \in X$ there is exactly one geodesic segment $j\colon [0, d(x,y)] \to X$ such that $j(0) = x$ and $j(d(x,y)) = y$.

Proposition 0.1. *Euclidean space is uniquely geodesic.*

Proof. If g is a geodesic segment that connects x to y and z is a point in the image of g, then, by definition, $\|x-z\| + \|z-y\| = \|x-y\|$. But, using the Cauchy inequality from linear algebra, it is easy to see that the Euclidean triangle inequality becomes an equality if and only if z is in the image of the linear segment $j|_{[0,\|x-y\|]}$. □

If a metric space X is uniquely geodesic and $x, y \in X$, $x \neq y$, we denote the image of the unique geodesic segment connecting x to y by $[x, y]$.

A *triangle* in Euclidean space consists of three points $A, B, C \in \mathbb{E}^n$ (the *vertices*) and of the three *sides* $[A, B]$, $[B, C]$ and $[C, A]$. Let the lengths of the sides be, in the corresponding order, c, a and b, and let the angles between the sides at the vertices A, B and C be α, β and γ. These quantities are connected via the

Euclidean law of cosines.

$$c^2 = a^2 + b^2 - 2ab\cos\gamma\ .$$

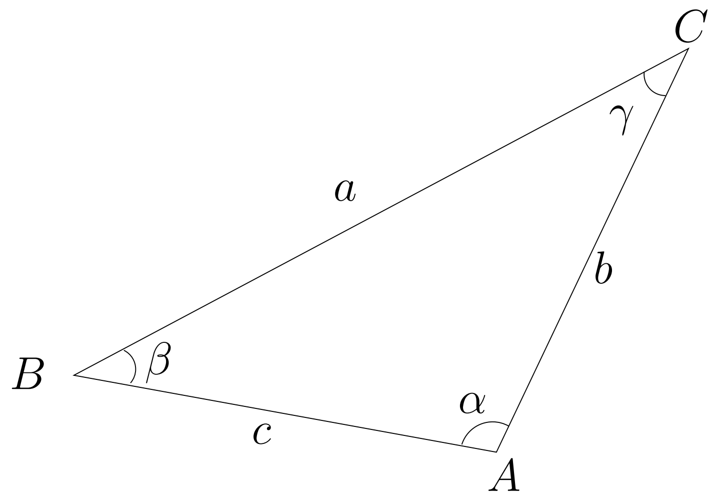

FIGURE 1

Proof. The proof is linear algebra:

$$\begin{aligned} c^2 = \|B-A\|^2 = \|B-C+C-A\|^2 &= b^2 + 2(B-C|C-A) + a^2 \\ = b^2 + 2(B-C|C-A) + a^2 &= b^2 - 2ab\cos\gamma + a^2\ . \end{aligned}$$

□

The law of cosines can be proved without knowing that $\mathbb{E}^n$ is uniquely geodesic. In fact, using the law of cosines, it is easy to prove that Euclidean space is uniquely geodesic, compare with the case of the sphere treated in section 0.3 .

0.3. **The sphere.** The unit sphere in $(n-1)$-dimensional Euclidean space is

$$\mathbb{S}^n = \{x \in \mathbb{E}^{n+1} : \|x\| = 1\} .$$

Let us show that the *angle distance*

$$d(x,y) = \arccos(x|y) \in [0,\pi]$$

is a metric. In order to do this, we will use the analog of the Euclidean law of cosines, but first we have to define the objects that are studied in spherical geometry.

Each 2-dimensional linear subspace (plane) $T \subset \mathbb{R}^{n+1}$ intersects $\mathbb{S}^n$ in a *great circle.* If $A \in \mathbb{S}^n$ and $u \in \mathbb{S}^n$ is orthogonal to A ($u \in A^\perp$), then the path $j_{A,u} \colon \mathbb{R} \to \mathbb{S}^n$,

$$j_{A,u}(t) = A\cos t + u \sin t ,$$

parametrises the great circle $\langle A, u\rangle \cap \mathbb{S}^n$, where $\langle A, u\rangle$ is the linear span of A and u. The vectors A and u are linearly independent, so $\langle A, u\rangle$ is a 2-plane.

If $A, B \in \mathbb{S}^n$ such that $B \neq \pm A$, then there is a unique plane that contains both points. Thus, there is unique great circle that contains A and B, in the remaining cases, there are infinitely many such planes. The great circle is parametrised by the map $j_{A,u}$, with

$$u = \frac{B-(B|A)A}{\|B-(B|A)A\|} = \frac{B-(A|B)A}{\sqrt{1-(A|B)^2}} .$$

Now $j(0) = A$ and $j(d(A,B)) = B$.

If $B = -A$, then there are infinitely many great circles through A and B: the map $j_{A,u}$ parametrises a great circle through A and B for any $u \in A^\perp$.

We call the restriction of any $j_{A,u}$ as above to any compact interval $[0,s]$ a *spherical segment*, and u is called the *direction* of $j_{A,u}$. Once we have proved that d is a metric, it is immediate that a spherical segment is a geodesic segment.

A triangle in $\mathbb{S}^n$ is defined as in the Euclidean case but now the sides of the triangle are the spherical segments connecting the vertices.

Let $j_{A,u}([0,d(C,A)])$ be the side between C and A, and let $j_{A,v}([0,d(C,B)])v$ be the side between C and B. The angle between $j_{A,u}([0,d(C,A)])$ and $j_{A,v}([0,d(C,B)])$ is $\arccos(u|v)$, which is the angle at A between the sides $j_{A,u}([0,d(A,B)])$ and $j_{A,v}([0,d(A,B)])$ in the ambient space $\mathbb{E}^{n+1}$. Now we can state and prove the

Spherical law of cosines.

$$\cos c = \cos a \cos b + \sin a \sin b \cos\gamma .$$

Proof. Let u and v be the initial tangent vectors of the hyperbolic segments $j_{C,u}$ from C to A and $j_{C,v}$ from C to B. As u and v are orthogonal to C, we have

$$\begin{aligned}\cos c = (A|B) &= (\cos(b)C + \sin(b)u|\cos(a)C + \sin(a)v)\\ &= \cos(a)\cos(b) + \sin(b)\sin(a)(u|v) .\end{aligned}$$

□

Proposition 0.2. *The angle metric is a metric on* $\mathbb{S}^n$*.* $(\mathbb{S}^n, d)$ *is a geodesic metric space. If* $d(A,B) < \pi$*, then there is a unique geodesic segment from* A *to* B*.*

Proof. Clearly, the triangle inequality is the only property that needs to be checked to show that the angle metric is a metric. Let $A, B, C \in \mathbb{S}^n$ be three distinct points and use the notation introduced above for triangles. The function

$$\gamma \mapsto f(\gamma) = \cos(a)\cos(b) + \sin(a)\sin(b)\cos(\gamma)$$

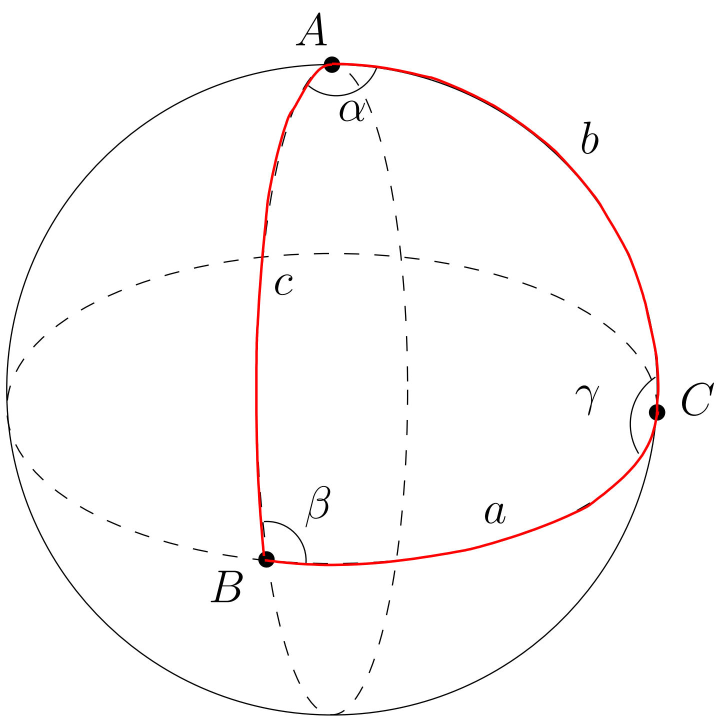

FIGURE 2

is strictly decreasing on the interval $[0, \pi]$, and

$$f(\pi) = \cos(a)\cos(b) - \sin(b)\sin(a) = \cos(a+b)\,.$$

Thus, the law of cosines implies that for all $\gamma \in [0, \pi]$, we have

$$\cos(c) = \cos(a)\cos(b) + \sin(a)\sin(b)\cos(\gamma) \geq \cos(a+b),$$

which implies $c \leq a + b$. Thus, the angle metric is a metric.

Equality holds in the triangle inequality if and only if $\gamma = \pi$. In this case, all the points A, B and C lie on the same great circle and C is contained in the side connecting A to B. Thus, the spherical segments are the only geodesic segments connecting A and B.

If $A \neq \pm B$, then there is exactly one 2-plane containing both points. The points A and B divide the great circle containing them in two parts on unequal lengths. This proves the third claim. □

Note that the sphere has no geodesic lines or rays because the diameter of the sphere is π.

1. Hyperbolic geometry

In this section, we define define hyperbolic space using the hyperboloid model which is analogous to the sphere that was treated in the warmup section.

1.1. **Minkowski space.** In this section, we need some basic facts on bilinear forms, which can be studied for example from [Gre]. Let V and W be real vector spaces. A map $\Phi\colon V \times W \to \mathbb{R}$ is a *bilinear form*, if the maps $v \mapsto \Phi(v, w_0)$ and $v \mapsto \Phi(v_0, w)$ are linear for all $w_0 \in W$ and all $v_0 \in V$. A bilinear form Φ is *nondegenerate* if

- $\Phi(x, y) = 0$ for all $y \in W$ only if $x = 0$, and
- $\Phi(x, y) = 0$ for all $x \in V$ only if $y = 0$.

If $W = V$, then Φ is *symmetric* if $\Phi(x, y) = \Phi(y, x)$ for all $x, y \in V$. It is

- *positive semidefinite* if $\Phi(x, x) \geq 0$ for all $x \in V$,

- *positive definite* if $\Phi(x,x) > 0$ for all $x \in V - \{0\}$,
- *negative (semi)definite* if $-\Pi$ is positive (semi)definite, and
- *indefinite* otherwise.

The quadratic form corresponding to a bilinear form $\Phi\colon V \times V \to \mathbb{R}$ is the function $q\colon V \to \mathbb{R}$, $q(x) = \Phi(x,x)$. A positive definite symmetric bilinear form is an inner product.

If V is a vector space with a symmetric bilinear form Φ, we say that two vectors $u, v \in V$ are *orthogonal* if $\Phi(u,v) = 0$, and this is denoted as usual by $u \perp v$. The *orthogonal complement* of $u \in V$ is

$$u^\perp = \{v \in V : u \perp v\}.$$

Let us consider the indefinite nondegenerate symmetric bilinear form $\langle \cdot | \cdot \rangle$ on $\mathbb{R}^{n+1}$ given by

$$\langle x|y \rangle = -x_0 y_0 + \sum_{i=1}^{n} x_i y_i,$$

where $x = \{(x_0, x_1, \dots, x_n)$. We call $\langle \cdot | \cdot \rangle$ the *Minkowski bilinear form*, and the pair $\mathbb{R}^{1,n} = (\mathbb{R}^{n+1}, \langle \cdot | \cdot \rangle)$ is the $n+1$-dimensional *Minkowski space*. A basis $\{v_0, v_1, \dots, v_n\}$ of $\mathbb{R}^{1,n}$ is *orthonormal* if the basis elements are pairwise orthogonal and if $\langle v_0 | v_0 \rangle = -1$ and $\langle v_i | v_i \rangle = 1$ for all $i \in \{1, 2, \dots, n\}$.

Minkowski space has a number of geometrically significant subsets: The variety

$$\mathscr{H}^n = \{x \in \mathbb{R}^{1,n} : \langle x|x \rangle = -1\}$$

is a two-sheeted hyperboloid, and its upper sheet is

$$\mathbb{H}^n = \{x \in \mathbb{R}^{1,:} \langle x|x \rangle = -1,\ v_0 > 0\}.$$

The subset of *null-vectors* is the *light cone*

$$\mathscr{L}^n = \{x \in \mathbb{R}^{1,n} : \langle x|x \rangle = 0\}\,.$$

The name light cone comes from Einstein's special theory of relativity, which lives in $\mathbb{R}^{1,3}$. Furthermore, we will occasionally use the physical terminology and say that a vector is

- *lightlike* if $\langle x|x \rangle = 0$,
- *timelike* if $\langle x|x \rangle < 0$, and
- *spacelike* if $\langle x|x \rangle > 0$.

1.2. **Hyperbolic space.** The metric space $(\mathbb{H}^n, d)$, where

$$d(x,y) = \operatorname{arcosh}(-\langle x|y \rangle) \in [0, \infty[\,,$$

is the *hyperboloid model* of n-dimensional *(real) hyperbolic space*. The metric d is the *hyperbolic metric*.

In fact, we still need to show that the hyperbolic metric is a metric. The proof follows the same idea that was used to treat the angle metric for the sphere $\mathbb{S}^n$.

Lemma 1.1. *If $u, v \in \mathbb{H}^n$, then $\langle u|v \rangle \leq -1$ with equality only if $u = v$.*

Proof. Using the Cauchy inequality for the Euclidean inner product in $\mathbb{R}^n$ for the first inequality and a simple calculation for the second, we have

$$\begin{aligned}\langle u|v\rangle &= -u_0v_0 + \sum_{i=1}^n u_iv_i \leq -u_0v_0 + \sqrt{\sum_{i=1}^n u_i^2}\sqrt{\sum_{i=1}^n v_i^2} \\ &= -u_0v_0 + \sqrt{u_0^2-1}\sqrt{u_0^2-1} \leq -1\,.\end{aligned}$$

Any line through the origin intersects $\mathbb{H}^n$ in at most one point, so Cauchy's inequality is an equality if and only if $u = v$. □

We will also need the following standard result for nondegenerate bilinear forms of signature $(1, n)$

Lemma 1.2. *For any $u \in \mathbb{H}^n$, the restriction of the Minkowski bilinear form to $a^\perp$ is positive definite.* □

Proof. This follows from Sylvester's law of inertia, see for example [Gre, Chapter IX]. □

Let $a \in \mathbb{H}^n$, and let $u \in a^\perp$ such that $\langle u|u\rangle = 1$. The mapping $j_{a,u} \colon \mathbb{R} \to \mathbb{H}^n$,

$$j_{a,u}(t) = a\cosh(t) + u\sinh(t)\,,$$

is the *hyperbolic line* through a in *direction* u. It is easy to check that, indeed, the image of $j_{a,u}$ is contained in $\mathbb{H}^n$ and that for all $s, t \in \mathbb{R}$, we have

$$d(j_{a,u}(t), j_{a,u}(s)) = |s-t|\,. \tag{1}$$

As in section 0.3 for the sphere, if we show that d is a metric, then $j_{a,u}$ is a geodesic line. We define hyperbolic segments and rays as the appropriate restrictions of the geodesic line.

Lemma 1.3. *For any $a \in \mathbb{H}^n$ and any $u \in a^\perp$, $j_{a,u}(\mathbb{R}) = \mathbb{H}^n \cap \langle a, u\rangle$. If a 2-plane T intersects $\mathbb{H}^n$, then $T \cap \mathbb{H}^n$ is the image of a hyperbolic line.*

Proof. Clearly, the image of $j_{a,u}$ is contained in the 2-plane $\langle a, u\rangle$.

On the other hand, if a plane $T = \langle u, v\rangle$ intersects $\mathbb{H}^n$ at two distinct points p and q, the geodesic line $j_{p,u}$ with

$$u = \frac{q + \langle p|q\rangle p}{|q + \langle p|q\rangle p|}$$

passes through p and q. If we fix $p \in \mathbb{H}^n$, there are exactly two unit tangent vectors v and $-v$ in $T_p\mathbb{H}^n \cap T$, and the hyperbolic lines $j_{p,v}$ and $j_{p,-v}$ defined by these vectors have the same image. Therefore, all points in $\mathbb{H}^n \cap T$ are contained in $j_{p,v}(\mathbb{R})$ for any $p \in \mathbb{H}^n$. □

Lemma 1.4. *For any $a \in \mathbb{H}^n$, the tangent space $T_a\mathbb{H}^n$ of $\mathbb{H}^n$ at a coincides with $a^\perp$.*

Proof. The orthogonal complement $a^\perp$ has dimension n because the Minkowski bilinear form is nondegenerate. Each vector in $a^\perp$ is the tangent vector at a of a smooth curve contained in $\mathbb{H}^n$. □

We will not need this result at this point, but it is good to observe that the restriction of the Minkowski bilinear form to each tangent space defines a Riemannian metric.

We define the *angle* $\measuredangle(u,v)$ of any two vectors $u, v \in T_a\mathbb{H}^n = a^\perp - \{0\}$, using the inner product induced from the Minkowski bilinear form:

$$\measuredangle(u,v) = \arccos(\langle u|v\rangle)$$

The inner product induces a norm

$$|u| = \sqrt{\langle u|u\rangle}$$

on $a^\perp$ for all $a \in \mathbb{H}^n$.

The first hyperbolic law of cosines.

$$\cosh c = \cosh a \cosh b - \sinh a \sinh b \cos\gamma\ .$$

Proof. Let u and v be the initial tangent vectors of the hyperbolic segments from C to A and from C to B. As u and v are orthogonal to C, we have as in the spherical case,

$$\begin{aligned}\cosh c &= -\langle A|B\rangle = -\langle \cosh(b)C + \sinh(b)u|\cosh(a)C + \sinh(a)v\rangle \\ &= \cosh(a)\cosh(b) - \sinh(b)\sinh(a)\langle u|v\rangle\ .\end{aligned}$$ □

Theorem 1.5. *Hyperbolic space is a uniquely geodesic metric space. Hyperbolic lines, rays and segments are geodesic lines, rays and segments.*

Proof. The fact that the hyperbolic metric is indeed a metric is proved in the same way as Proposition 0.2 in the spherical case. Now we consider the increasing function

$$\gamma \mapsto \cosh a \cosh b - \sinh a \sinh b \cos\gamma,$$

which attains its maximum value $\cosh(a+b)$ when $\gamma = \pi$. The claim on hyperbolic lines, rays and segments follows from equation (1).

If p and q are distinct points in $\mathbb{H}^n$, there is a unique 2-plane through them. Thus, there is exactly one hyperbolic line through these points. As in the spherical case, we see that the triangle inequality in hyperbolic geometry is an equality if and only if the third point z lies in the hyperbolic segment between x and y. □

The law of cosines implies that a triangle in $\mathbb{E}^n$, $\mathbb{S}^n$ or $\mathbb{H}^n$ is uniquely determined up to an isometry of the space, if the lengths of the three sides are known. In Euclidean space, the three angles of a triangle do not determine the triangle uniquely. In $\mathbb{S}^n$ and $\mathbb{H}^n$ the angles determine a triangle uniquely. For $\mathbb{H}^n$, this is the content of

The second hyperbolic law of cosines.

$$\cosh c = \frac{\cos\alpha\cos\beta + \cos\gamma}{\sin\alpha\sin\beta}\ .$$

This formula follows from the first law of cosines by a lengthy manipulation, see [Bea, p. 148–150].

Recall that in Euclidean space, the law of sines states that the identity

$$\frac{a}{\sin\alpha} = \frac{b}{\sin\beta} = \frac{c}{\sin\gamma}$$

holds for any triangle. An analogous result holds in hyperbolic space:

The hyperbolic law of sines.

$$\frac{\sinh a}{\sin\alpha} = \frac{\sinh b}{\sin\beta} = \frac{\sinh c}{\sin\gamma}\ .$$

Proof. The first law of cosines implies that

$$\left(\frac{\sinh c}{\sin\gamma}\right)^2 = \frac{\sinh^2 a \sinh^2 b \sinh^2 c}{2\cosh a\cosh b\cosh c - \cosh^2 a - \cosh^2 b - \cosh^2 c + 1}\,.$$

The claim follows because this expression is symmetric in a, b and c. □

1.3. **Isometries.** We will now study the isometries of hyperbolic space, that is, mappings of hyperbolic space to itself that preserve distances. We begin by introducing some convenient terminology: A group G *acts* on a metric space X *by isometries* if there is a homomorphism $\Phi\colon G \to \mathrm{Isom}(X)$. Normally, one ignores the homomorphism Φ in notation, and writes $g(x)$ or $g \cdot x$ or something similar to mean $\Phi(g)(x)$. Similarly, one defines action by homomorphisms in a topological space, a linear action in a vector space etc.

If a group G acts on a space X, and x is a point in X, the set

$$G(x) = \{g(x) : g \in g\}$$

is the *G-orbit* of x. The action of a group is said to be *transitive* if $G(x) = X$ for some (and therefore for any) $x \in X$. For any nonempty subset A of X, the *stabiliser* of A in G is

$$\mathrm{Stab}_G A = \{g \in G : gA = A\},.$$

Clearly, stabilisers are subgroups of G.

It is good to remember some facts from Euclidean and spherical geometry: The *(Euclidean) orthogonal group* of dimension n is

$$\begin{aligned}\mathrm{O}(n) &= \{A \in \mathrm{GL}_n(\mathbb{R}) : (Ax|Ay) = (x|x) \text{ for all } x, y \in \mathbb{E}^n\} \\ &= \{A \in \mathrm{GL}_n(\mathbb{R}) : {}^T\!AA = I_n\}\,.\end{aligned}$$

It is easy to check that if $A \in \mathrm{O}(n)$ and $b \in \mathbb{R}^n$, then the mapping $x \mapsto Ax + b$ is an isometry of $\mathbb{E}^n$, and that the mapping $x \mapsto Ax$ is an isometry of $\mathbb{S}^{n-1}$.

We will now introduce the corresponding group for $\mathbb{H}^n$. Let $J_{1,n} = \mathrm{diag}(-1, 1, \ldots, 1)$, and note that

$$\langle x|y\rangle = {}^T\!xJy$$

for all $x, y \in \mathbb{H}^n$. The *orthogonal group* of the Minkowski bilinear form is

$$\begin{aligned}\mathrm{O}(1, n) &= \{A \in \mathrm{GL}_n(\mathbb{R}) : \langle Ax|Ay\rangle = \langle x|y\rangle \text{ for all } x, y \in \mathbb{R}^{1,n}\} \\ &= \{A \in \mathrm{GL}_n(\mathbb{R}) : {}^T\!AJ_{1,n}A = J_{1,n}\}\,.\end{aligned}$$

Clearly, the linear action of $\mathrm{O}(1, n)$ on $\mathbb{R}^{1,n}$ preserves the two-sheeted hyperboloid $\mathscr{H}^n$.

Let us write an $(n + 1) \times (n + 1)$-matrix A in terms of its column vectors $A = (a_0, a_1, \ldots, a_n)$. If $A \in \mathrm{O}(1, n)$, then $a_0 = A(x_0)$ for $x_0 = 1, 0, \ldots, 0) \in \mathbb{H}^n$. Thus $A(x_0) \in \mathbb{H}^n$ if and only if $A_{00} > 0$, and therefore the stabiliser in $\mathrm{O}(1, n)$ of the upper sheet $\mathbb{H}^n$ is

$$\begin{aligned}\mathrm{O}^+(1, n) &= \{A \in O(1, n) : A\mathbb{H}^n = \mathbb{H}^n\} \\ &= \{A \in \mathrm{GL}_n(\mathbb{R}) : A_{00} > 0,\ \langle Ax|Ay\rangle = \langle x|y\rangle \text{ for all } x, y \in \mathbb{R}^{1,n}\} \\ &= \{A \in \mathrm{GL}_n(\mathbb{R}) : A_{00} > 0,\ {}^T\!AJ_{1,n}A = J_{1,n}\}\,,\end{aligned}$$

which is the identity component of $\mathrm{O}(1, n)$.

The following observation is proved in the same way as its Euclidean analog:

Lemma 1.6. *An $(n+1)\times(n+1)$-matrix $A=(a_0,a_1,\dots,a_n)$ is in $\mathrm{O}(1,n)$ if and only if the vectors $a_0,a_1,\dots,a_n$ form an orthonormal basis of $\mathbb{R}^{1,n}$. Furthermore, $A\in\mathrm{O}^+(1,n)$ if and only if $A\in\mathrm{O}(1,n)$ and $a_0\in\mathbb{H}^n$.* □

Using Lemma 1.6, it is not difficult to show that the following analog of the Euclidean and spherical cases holds in hyperbolic space:

Proposition 1.7. $\mathrm{O}^+(1,n)$ *acts transitively by isometries on $\mathbb{H}^n$. In particular,* $\mathrm{Isom}(\mathbb{H}^n)$ *acts transitively on $\mathbb{H}^n$.*

Proof. Let $g\in\mathrm{O}^+(1,n)$, and let $x,y\in\mathbb{H}^n$. By the definition of the hyperbolic metric and of $\mathrm{O}^+(1,n)$, we have

$$d(g(x),g(y))=\operatorname{arcosh}(-\langle g(x)|g(y)\rangle=\operatorname{arcosh}(-\langle x|y\rangle=d(x,y)\,.$$

Transitivity follows from the fact that any orthonormal basis of $\mathbb{R}^{1,n}$ whose first vector is in $\mathbb{H}^n$ can be mapped to any other similar one by a transformation in $\mathrm{O}^+(1,n)$: If $p\in\mathbb{H}^n$, and the vectors $v_1,v_2,\dots,v_n$ form an orthogonal basis of $p^\perp=T_p\mathbb{H}^n$, then the matrix $A=(p,v_1,\dots,v_n)\in\mathrm{O}(1,n)$ gives an isometry which maps $(1,0,\dots,0)$ to p. □

Example 1.8. Let $t\in\mathbb{R}$. The matrix

$$L_t=\begin{pmatrix}\cosh t & \sinh t & 0\\ \sinh t & \cosh t & 0\\ 0 & 0 & 1\end{pmatrix}\in\mathrm{O}(1,2) \tag{2}$$

acts on $\mathbb{H}^2$ as an isometry that preserves the intersection of $\mathbb{H}^2$ with any 2-plane $\{x\in\mathbb{R}^{1,2}:x_2=c\}$, in particular, it stabilises the geodesic line

$$\ell=\{x\in\mathbb{H}^3:x_2=0\}\,.$$

For any point $p=(a,b,0)\in\ell$, we have

$$d(L_t(p),p)=\operatorname{arcosh}(-\langle L_tp|p\rangle)=\operatorname{arcosh}((-a^2+b^2)\cosh(t))=|t|\,.$$

For any $\theta\in\mathbb{R}$, let $\widehat{R}_\theta=\begin{pmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{pmatrix}\in\mathrm{O}(2)$, and let

$$R_\theta=\operatorname{diag}(1,\widehat{R}_\theta)=\begin{pmatrix}1 & 0\\ 0 & \widehat{R}(\theta)\end{pmatrix}=\begin{pmatrix}1 & 0 & 0\\ 0 & \cos\theta & -\sin\theta\\ 0 & \sin\theta & \cos\theta\end{pmatrix}\in\mathrm{O}(1,2)\,. \tag{3}$$

This mapping rotates the hyperboloid around the vertical axis by the angle θ. Another important mapping that comes by extension from $\mathrm{O}(2)$ is given by the matrix $\operatorname{diag}(1,1,-1)$, which is a reflection in the geodesic line ℓ defined above.

For each $v\in\mathscr{L}^2$ and $c<0$, the set

$$\{x\in\mathbb{H}^2:\langle v|x\rangle=c\}$$

is called a *horosphere* based at v. The mapping given by the matrix

$$N_s=\begin{pmatrix}1+\frac{s^2}{2} & -\frac{s^2}{2} & s\\ \frac{s^2}{2} & 1-\frac{s^2}{2} & s\\ s & -s & 1\end{pmatrix}\in\mathrm{O}(1,2) \tag{4}$$

maps each horosphere based at $(1,1,0) \in \mathscr{L}^2$ to itself.

———

Composing some number of the above mappings we obtain further examples of isometries of the hyperbolic plane. For example, if $p \in \mathbb{H}^2$, then there is some $\theta \in \mathbb{R}$ such that $R_\theta(p) \in \ell$. Now, $L^{-1}_{d(o,p)}(R_\theta(p)) = L_{-d(o,p)}(R_\theta(p)) = (1,0,0)$, and for any $\phi \in \mathbb{R}$, the mapping $S = R_{-\theta} \circ L_{d(o,p)} \circ R_\phi \circ L^{-1}_{d(o,p)} \circ R_\theta$ is an isometry that fixes p and maps each sphere centered at p to itself. The mapping S is *conjugate* to R_ϕ in $\mathrm{Isom}(\mathbb{H}^n)$.

The isometries introduced above are classified according to the conic sections they correspond to. The mapping L_t and any of its conjugates in $\mathrm{Isom}(\mathbb{H}^n)$ is called *hyperbolic* because L_t maps each affine plane parallel to the (x_0, x_1)-plane in $\mathbb{R}^{1,2}$ to itself, and these planes intersect the lightcone in hyperbola, which is degenerate for the (x_0, x_1)-plane itself.

The mapping $R(\theta)$ and any of its conjugates is called *elliptic* because it preserves all horizontal hyperplanes in $\mathbb{R}^{1,2}$ and their intersections with $\mathscr{L}^2$, which are circles centered at $(1,0,0)$.

The mapping N_s and any of its conjugates is called *parabolic* because it preserves all affine hyperplanes $\{x \in \mathbb{R}^{1,2} : \langle v|x\rangle = c\}$, which intersect $\mathscr{L}^2$ in a parabola when $c < 0$.

All of these examples can be generalised to higher dimensions:

- L_t is extended as the identity on the last coordinates to $\mathrm{diag}(L_t, I_{n-2}) \in \mathrm{O}(1,n)$.
- Any Euclidean orthogonal matrix $A \in \mathrm{O}(n)$ gives an isometry $\mathrm{diag}(1, A) \in \mathrm{O}(1,n)$.
- N_s is extended as the identity on the last coordinates to $\mathrm{diag}(N_s, I_{n-2}) \in \mathrm{O}(1,n)$.

———

If T is an $(m+1)$-dimensional linear subspace of $\mathbb{R}^{n+1}$ that intersects $\mathbb{H}^n$, then $T \cap \mathbb{H}^n$ is an m-dimensional *hyperbolic subspace* of $\mathbb{H}^n$. If $m = n-1$, then T is a *hyperplane.* A modification of the proof of Proposition 1.7 gives

Proposition 1.9. *Any two hyperbolic subspaces of $\mathbb{H}^n$ can be mapped to each other by isometries of $\mathbb{H}^n$. In particular, a k-dimensional hyperbolic subspace of $\mathbb{H}^n$ is isometric to $\mathbb{H}^k$.* □

Any hyperplane T in $\mathbb{R}^{1,n}$ is of the form $T = u^\perp$ for some $u \in \mathbb{R}^{1,n} - \{0\}$ because the Minkowski bilinear form is nondegenerate. Let $H = u \perp \cap \mathbb{H}^n$ be a hyperbolic hyperplane. Since H intersects $\mathbb{H}^n$, it contains a vector v for which $\langle v|v\rangle = -1$. Sylvester's law implies that $\langle u|u\rangle > 0$, and after normalising, we may assume that u is a unit vector. The *reflection* in H is the map

$$r_H(x) = x - 2\langle x|u\rangle u\,.$$

Reflections are very useful isometries, the following results give some of their basic properties.

Proposition 1.10. *Let H be a hyperbolic hyperplane. Then*
(1) $r_H \circ r_H$ is the identity.
(2) $r_H \in O^+(1,n)$.
(3) $d(r_H(x), y) = d(x,y)$ for all $x \in \mathbb{H}^n$ and all $y \in H$.

(4) The fixed point set of r_H is H

Proof. (1) This easy computation is left as an exercise.
(2) Clearly, r_H is a linear mapping, and it is a bijection by (1). Using bilinearity and symmetry of the Minkowski form and the fact that u is a unit vector, we get

$$\begin{aligned}\langle r_H(x)|r_H(y)\rangle &= \langle x - 2\langle x|u\rangle u | y - 2\langle y|u\rangle u\rangle \\ &= \langle x|y\rangle - 2\langle y|u\rangle\langle x|u\rangle - 2\langle x|u\rangle\langle u|y\rangle + 4\langle x|u\rangle\langle y|u\rangle\langle u|u\rangle \\ &= \langle x|y\rangle\,.\end{aligned}$$

Thus, $r_H \in \mathrm{O}(1,n)$. Furthermore, for any $v \in H$,

$$r_H(v) = v - 2\langle v|u\rangle u = v,$$

so there are points in $\mathbb{H}^n$ which are mapped to $\mathbb{H}^n$. Since r_H is continuous and preserves the Minkowski form, $r_H(\mathbb{H}^n) \subset \mathbb{H}^n$, and therefore $r_H \in \mathrm{O}^+(1,n)$.
(3) For any $x \in \mathbb{H}^n$ and all $y \in H$, we have

$$\langle r_H(x)|y\rangle = \langle x - 2\langle x|u\rangle u|y\rangle = \langle x|y\rangle - 2\langle x|u\rangle\langle u|y\rangle = \langle x|y\rangle,$$

where the final equality follows from the assumption $u \in H^\perp$.
(4) This follows immediately from (2) by taking $x = y \in H$. □

The *bisector* of two distinct points p and q in $\mathbb{H}^n$ is the hyperplane

$$\mathrm{bis}(p,q) = \{x \in \mathbb{H}^n : d(x,p) = d(x,q)\} = (p-q)^\perp \cap \mathbb{H}^n.$$

Proposition 1.11. *(1) For any $p, q \in \mathbb{H}^n$, the bisector $\mathrm{bis}(p,q)$ is a hyperbolic hyperplane.*
(2) If $r_H(x) = y$ and $x \notin H$, then $H = \mathrm{bis}(x,y)$.
(3) If $p, q \in \mathbb{H}^n$, $p \neq q$, then $r_{\mathrm{bis}(p,q)}(p) = q$.
(4) Let $\phi \in \mathrm{Isom}(\mathbb{H}^n)$, $\phi \neq \mathrm{id}$. If $a \in \mathbb{H}^n$, $\phi(a) \neq a$, then the fixed points of ϕ are contained in $\mathrm{bis}(a, \phi(a))$.
(5) Let $\phi \in \mathrm{Isom}(\mathbb{H}^n)$, $\phi \neq \mathrm{id}$. If H is a hyperplane such that $\phi|_H$ is the identity, then $\phi = r_H$.

Proof. (1) Since $\langle p-q|p-q\rangle = -2 - 2\langle p|q\rangle > 0$, $(p-q)^\perp$ contains a vector v for which $\langle v|v\rangle < 0$, and therefore the hyperplane $(p-q)^\perp$ of $\mathbb{R}^{1,n}$ intersects $\mathbb{H}^n$.
(2) follows from Proposition 1.10(3).
(3) Now, $2\langle p|p-q\rangle = 2(\langle p|p\rangle - \langle p|q\rangle) = -2 - 2\langle p|q\rangle = |p-q|^2$. Thus,

$$r_{\mathrm{bis}(p,q)}(p) = p - 2\langle p|p-q\rangle\frac{p-q}{|p-q|^2} = q\,. \quad \square$$

(4) If $\phi(b) = b$, then $d(a,b) = d(\phi(a),\phi(b)) = d(\phi(a),b)$, so that $b \in \mathrm{bis}(a,\phi(a))$.
(5) is an instructive exercise.

Propositions 1.10 and 1.11 give a new proof of Proposition 1.7.

Next, we want to prove that all isometries of hyperbolic space are restrictions to $\mathbb{H}^n$ of linear automorphisms of $\mathbb{R}^{1,n}$:

Theorem 1.12. $\mathrm{Isom}(\mathbb{H}^n) = \mathrm{O}^+(1, n)$.

The idea of the proof is to show that each isometry of $\mathbb{H}^n$ is the composition of reflections in hyperbolic hyperplanes. In order to do this, we show that the isometry group has a stronger transitivity property than wwidehat was noted above.

Proposition 1.13. *Let $p_1, p_2, \dots, p_k, q_1, q_2, \dots, q_k \in \mathbb{H}^n$ be points that satisfy*

$$d(p_i, p_j) = d(q_i, q_j)$$

for all $i, j \in \{1, 2, \dots, k\}$. Then, there is an isometry $\phi \in \mathrm{Isom}(\mathbb{H}^n)$ such that $\phi(p_i) = q_i$ for all $i \in \{1, 2, \dots, k\}$. Furthermore, the isometry ϕ is the composition of at most k reflections in hyperplanes.

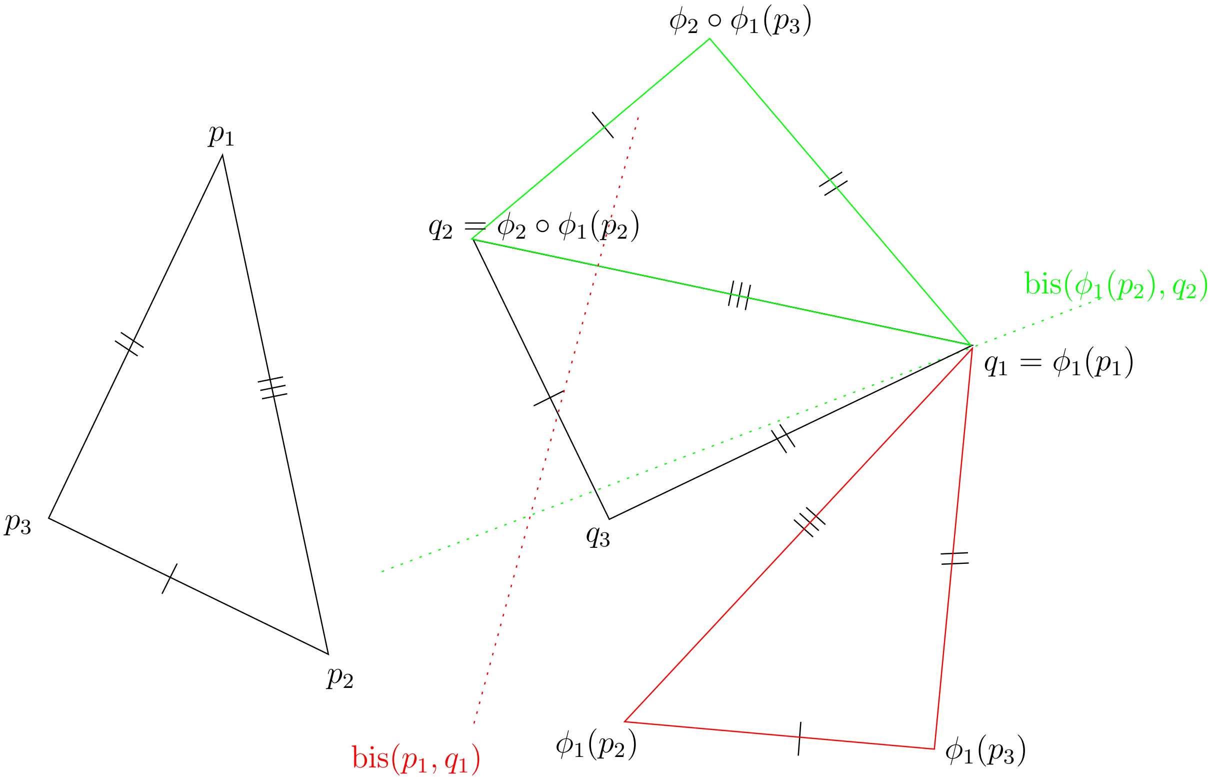

FIGURE 3

Proof. We construct the isometry by induction. If $p_1 = q_1$, let ϕ_1 be the identity, otherwise, let ϕ_1 be the reflection in the bisector of p_1 and q_1. Let $m > 1$ and assume that there is an isometry ϕ_m such that $\phi(p_i) = q_i$ for all $i \in \{1, 2, \dots, m\}$, which is the composition of at most m reflections.

Assume that $\phi_m(p_{m+1}) \neq q_{m+1}$. Now, $q_1, \dots q_m \in \mathrm{bis}(\phi_m(p_{m+1}), q_{m+1})$: For each $1 \le i \le m$, we have

$$d(q_i, \phi_m(p_{m+1})) = d(\phi_m(p_i), \phi_m(p_{m+1})) = d(p_i, p_{m+1}) = d(q_i, q_{m+1})\,.$$

The map $\phi_{m+1} = r_{\mathrm{bis}(\phi_m(p_{m+1}), q_{m+1})} \circ \phi_m$ satisfies $\phi_{m+1}(p_i) = q_i$ for all $1 \le i \le m+1$. □

Note that Proposition 1.13 implies that if T and T' are two triangles in $\mathbb{H}^n$ with equal angles or equal sides, then there is an isometry ϕ of $\mathbb{H}^n$ such that $\phi(T) = T'$.

Proof of Theorem 1.12. Let $\{a_0, a_1, \dots, a_n\}$ be a set of points in $\mathbb{H}^n$ which is not contained in any proper hyperbolic subspace. This is achieved by choosing them so that they generate $\mathbb{R}^{1,n}$ as a vector space. Proposition 1.13 implies that there is an isometry $\phi_0 \in \mathrm{O}^+(1,n)$ such that $\phi_0(\phi(a_1)) = a_i$ for all $1 \le i \le m+1$. Since the set of fixed points of $\phi_0 \circ \phi$ contains the points $a_0, a_1, \dots, a_{n+1}$, the fixed point set of ϕ_0 is not contained in a proper hyperbolic subspace. Proposition 1.11 implies that $\phi_0 \circ \phi$ is the identity map. Thus, $\phi = \phi_0^{-1}$. In particular, $\phi \in \mathrm{O}^+(1,n)$, which is all we needed to show. □

The same proof with "obvious" modifications works for Euclidean space and for the sphere. The corresponding results are that $\mathrm{Isom}(\mathbb{S}^n) = \mathrm{O}(n+1)$ and $\mathrm{Isom}(\mathbb{E}^n)$ is the group generated by $\mathrm{O}\, n$ and the translations $x \mapsto x + b$, $b \in \mathbb{R}^n$.

Corollary 1.14. *Any isometry of $\mathbb{H}^n$ can be represented as the composition of at most $n+1$ reflections.*

Proposition 1.15. *The stabiliser of any point $x \in \mathbb{H}^n$ is isomorphic to $\mathrm{O}(n)$.*

Proof. Consider first the point $x_0 = (1, 0, \dots, 0) \in \mathbb{H}^n$. If $A \in \mathrm{O}^+(1,n)$ fixes x_0, then the first column a_0 of A is $(1, 0, \dots, 0)$. As the other columns $a_1, \dots, a_n$ of A are orthogonal to a_0, we have $a_i = (0, x_i)$ with $x_i \in \mathbb{R}^n$ for all $1 \le i \le n$. The restriction of A to $x_0^\perp$ preserves the Euclidean inner product induced by the Minkowski form. Thus, $A = \mathrm{diag}(1, A_0)$, with $A_0 \in \mathrm{O}(n)$, and $\mathrm{Stab}_{\mathrm{Isom}(\mathbb{H}^n)}(x_0)$ is isomorphic to $\mathrm{O}(n)$.

For any $x \in \mathbb{H}$, let L_x be an isometry such that $L_x(x_0) = x$. The map $g \mapsto L_x \circ g \circ L_x^{-1}$ is an isomorphism between the groups $\mathrm{Stab}_{\mathrm{Isom}(\mathbb{H}^n)}(x_0)$ and $\mathrm{Stab}_{\mathrm{Isom}(\mathbb{H}^n)}(x)$. □

2. Models of hyperbolic space

In this section we consider a number of other models for hyperbolic space, that is, metric spaces (X, d_X) for which there is an isometry $h\colon (X, d_X) \to (\mathbb{H}^n, d)$. Hyperbolic space is the class of all metric spaces isometric with the hyperboloid model $(\mathbb{H}^n, d)$, and we can use any model that is best suited for the geometric problem at hand. After this section we will often talk about the "upper half plane model of $\mathbb{H}^n$" etc.

2.1. Klein's model.

Each line in $\mathbb{R}^{1,n}$ through the origin which intersects $\mathbb{H}^n$, intersects it in exactly one point, and it also intersects the embedded copy $\{1\} \times \mathbb{B}^n$ in $\mathbb{R}^{1,n}$ of the Euclidean n-dimensional unit ball $\mathbb{B}^n(0,1)$ in exactly one point. This correspondence determines a bijection $K\colon \mathbb{B}(0,1) \to \mathbb{H}^n$, which has the explicit expression

$$K(x) = \frac{(1,x)}{\sqrt{1 - \|x\|^2}}\,.$$

The map K becomes an isometry when we define a metric on $\mathbb{B}(0,1)$ by setting

$$d_K(x,y) = d(K(x), K(y)) = \operatorname{arcosh} \frac{1 - (x|y)}{\sqrt{1 - \|x\|^2}\sqrt{1 - \|y\|^2}}\,.$$

The metric space $(\mathbb{B}(0,1), d_K)$ is the *Klein model* of n-dimensional hyperbolic space.

Proposition 2.1. *As sets, the geodesic lines of the Klein model are Euclidean segments connecting two points in the Euclidean unit sphere.*

Proof. A geodesic line in $\mathbb{H}^n$ is the intersection of $\mathbb{H}^n$ with a 2-plane in $\mathbb{R}^{1,n}$. The intersection of this plane with $\mathbb{B}^n(0,1) \times \{1\}$ is the preimage under K of the geodesic line. □

The above observation makes it easy to show that the parallel axiom does not hold in $\mathbb{H}^n$.

2.2. **Poincaré's ball model.** Each affine line that passes through the point $(-1,0) \in \mathbb{R}\times\mathbb{R}^n = \mathbb{R}^{1,n}$ which intersects $\mathbb{H}^n$, intersects it in exactly one point, and it also intersects the n-dimensional ball $\{0\}\times B^n(0,1)$ embedded in $\mathbb{R}^{1,n}$ in exactly one point. This correspondence determines a bijection $P\colon \mathbb{B}(0,1)\to\mathbb{H}^n$,

$$P(x) = \Big(\frac{1+\|x\|^2}{1-\|x\|^2}, \frac{2x}{1-\|x\|^2}\Big)\,.$$

This expression is found by computing for any $x\in\mathbb{B}(0,1)$ that the point $y_t = (0,x)+t(1,x)$ on the line through the points $(0,x)$ and $(-1,0)$ of $\mathbb{R}\times\mathbb{R}^n=\mathbb{R}^{1,n}$ is in $\mathbb{H}^n$ if and only if $t=\frac{1+\|x\|^2}{1+\|x\|^2}$.

The map K becomes an isometry when we define a metric on $\mathbb{B}(0,1)$ by setting

$$d_P(x,y) = d(P(x),P(y)) = \operatorname{arcosh}\big(1+2\frac{\|x-y\|^2}{(1-\|x\|^2)(1-\|y\|^2)}\big)\,.$$

The metric space $(\mathbb{B}(0,1),d_P)$ is the *Klein model* of n-dimensional hyperbolic space.

Proposition 2.2. *As sets, the geodesic lines of the Poincaré model are the intersections of the Euclidean unit ball with Euclidean circles which are orthogonal to the unit sphere.*

Proof. The map $h=K^{-1}\circ P$ is an isometry between the Poincare and Klein models. A computation shows that

$$h(x) = \frac{2x}{1+\|x\|^2}.$$

(This can be done by observing that h is a radial map and then solving the equation

$$\frac{(1,y)}{\sqrt{1-y^2}} = \big(\frac{1+x^2}{1-x^2}, \frac{2x}{1-x^2}\big)$$

with $0\le x,y<1$.) On the other hand, the inversion S in the sphere centered at $(-1,0)$ of radius $\sqrt{2}$ has the expression

$$S(x) = \Big(\frac{1-\|x\|^2}{1+\|x\|^2}, \frac{2x}{1+\|x\|^2}\Big)\,,$$

so that if $\mathrm{pr}\colon\mathbb{E}^{n+1}\to\mathbb{E}^n$ is the Euclidean orthogonal projection, we have $h=\mathrm{pr}\circ S$. See Appendix **??** for a short review of some basic properties of inversions.

Note that $\{0\}\times\mathbb{S}(0,1)$ is contained in the fixed sphere of S. The inversion S maps any circle orthogonal to $\{0\}\times\mathbb{S}(0,1)$ to a circle on the unit sphere in $\mathbb{E}^{n+1}$ orthogonal to $\{0\}\times\mathbb{S}(0,1)$. These circles are the intersections of the unit sphere with 2-planes parallel to the x_0-axis, and thus, pr maps them to the geodesic lines of the Klein model. As h is an isometry, the result follows. □

Note that the mapping S in the proof of the above result is (the inverse) of the stereographic projection.

2.3. **The upper halfspace model.** Let

$$\mathbb{U}^n = \{x \in \mathbb{R}^n : x_n > 0\}$$

be the n-dimensional *upper halfspace*. Let S be the inversion in the sphere of center $(0,-1) \in \mathbb{R}^{n-1} \times \mathbb{R} = \mathbb{E}^n$ of radius $\sqrt{2}$. Now, the map $F = S|_{\mathbb{B}(0,1)} \colon \mathbb{B}(0,1) \to \mathbb{U}^n$ is a bijection, which becomes an isometry if we use the metric

$$d_{\mathbb{U}}(x,y) = d_P(F^{-1}(x), F^{-1}(y)) = \operatorname{arcosh}(1 + \frac{\|x-y\|^2}{2x_n y_n}) \tag{5}$$

in $\mathbb{U}^n$. The metric space $(\mathbb{U}^n, d_{\mathbb{U}})$ is the *upper halfspace model* of n-dimensional hyperbolic space.

Proposition 2.3. *As sets, the geodesic lines of the upper halfspace model are the intersections of* $\mathbb{U}^n$ *with Euclidean circles and lines which are orthogonal to* $\mathbb{E}^{n-1} \times \{0\}$.

Proof. The inversion used in the definition of the upper halfspace model maps lines and circles to lines or circles and preserves angles. See Appendix **??** for the basic properties of inversions. □

In practical applications, it is good to remember that a circle is perpendicular to $\mathbb{E} \times \{0\} \subset \mathbb{E}^2$ if and only if its center is in $\mathbb{E} \times \{0\}$.

2.4. **Riemannian metrics.** The restriction of the Minkowski bilinear form $\langle \cdot | \cdot \rangle$ to the tangent space $T_p\mathbb{H}^n = p^\perp$ of any point in the hyperboloid model is positive definite, and it defines a Riemannian metric on the hyperboloid.

The Riemannian length of a piecewise smooth path $\gamma \colon [a,b] \to \mathbb{H}^n$ is

$$\ell(\gamma) = \int_a^b \sqrt{\langle \dot\gamma(t) | \dot\gamma(t) \rangle}\, dt\,.$$

The length metric of the Riemannian metric of $\mathbb{H}^n$ is

$$d_{\text{Riem}}(x,y) = \inf \ell(\gamma),$$

where the infimum is taken over all piecewise smooth paths that connect x to y.

By definition of the Riemannian metric as the restriction of the Minkowski bilinear form to each tangent space, the group $\mathrm{O}(1,n)$ acts by Riemannian isometries on the hyperboloid. Thus, it is not surprising that the following result holds:

Proposition 2.4. *The length metric of the Riemannian metric of hyperbolic space is the hyperbolic metric.*

Let us prepare for the proof with some technical results:

Lemma 2.5. *Any sphere of radius* r *in* $\mathbb{H}^n$ *is isometric with the Euclidean sphere of radius* $\sinh r$ *if we use the Riemannian metric.*

Proof. As everything is invariant under isometries, it suffices to consider the sphere centered at $x = (1, 0, \ldots, 0)$. If $y \in \mathbb{S}(x,r)$, then $\cosh r = -\langle y | x \rangle = y_0$. Thus,

$$\mathbb{S}(x,r) = \{(y_0, \bar y) \in \mathbb{R}^{1,n} : y_0 = \cosh r,\ \|\bar y\| = \sinh r\} = (\{\cosh r\} \times \mathbb{E}^n) \cap \mathbb{H}^n\,.$$

This means that the tangent space of $\mathbb{S}(x,r)$ is contained in $\{\cosh r\} \times \mathbb{E}^n$, which implies that for any $p \in \mathbb{S}(x,r)$, the restriction of the Minkowski bilinear form to $T_p\mathbb{S}(x,r) \subset T_p\mathbb{H}^n$ is the standard Euclidean inner product. This proves the result. □

Lemma 2.6. *For any* $p \in \mathbb{H}^n$, *any geodesic ray starting from* p *intersects any sphere centered at* p *perpendicularly.*

Proof. Exercise. □

Let $x \in \mathbb{H}^n$, and let (r, u) be spherical coordinates for $T_x\mathbb{H}^n$ with $r \geq 0$ and u a unit vector. The *exponential map* $\exp_x : T_x\mathbb{H}^n \to \mathbb{H}^n$,

$$\exp_x(ru) = x \cosh r + u \sinh r = j_{x,u}(r)$$

is a bijection because $\mathbb{H}^n$ is uniquely geodesic. We call the spherical coordinates of $T_x\mathbb{H}^n$ the spherical coordinates of $\mathbb{H}^n$ centered at x. Note that this agrees with the usual definition of the exponential map in Riemannian geometry.

Lemma 2.7. *The Riemannian metric of* $\mathbb{H}^n$ *in spherical coordinates is*

$$dr^2 + \sinh^2(r)\, du^2 .$$

Proof. Apply the above two lemmas. □

Proof of Proposition 2.4. Let $x, y \in \mathbb{H}^n$, and let $\gamma : [0, 1] \to \mathbb{H}^n$ be a path that connects x to y. In spherical coordinates centered at x, $\gamma = r(t)u(t)$, so Lemma 2.7 gives

$$\langle \dot\gamma(t) | \dot\gamma(t) \rangle = \dot r(t)^2 + \sinh^2(t) \langle \dot u(t) | \dot u(t) \rangle,$$

and thus, the length of the curve γ is

$$\begin{aligned}\ell(\gamma) &= \int_0^1 \sqrt{\langle \dot\gamma(t) | \dot\gamma(t) \rangle \, dt} = \int_0^1 \sqrt{\dot r(t)^2 + \sinh^2 r(t) \langle \dot u(t) | \dot u(t) \rangle \, dt} \\ &\geq \int_0^1 |\dot r(t)| \, dt \geq r\end{aligned}$$

that is, at least the length of the hyperbolic geodesic segment with the same endpoints. □

Proposition 2.4 allows us to use the Riemannian structure of hyperbolic space in any of the models introduced above. The expressions in the Poincaré and upper halfspace models are particularly useful. The proof of the following result is a straight-forward computation.

Proposition 2.8. *(1) The Riemannian metric of the ball model is* $\frac{4(\cdot|\cdot)}{(1-\|x\|^2)^2}$.

(2) The Riemannian metric of the upper halfspace model is $\frac{(\cdot|\cdot)}{x_n^2}$.

Proof. (1) For all tangent vectors u, v in $T_xB(0, 1)$, we have

$$\langle DP(x)u | DP(x)v \rangle = \frac{4(u|v)}{(1 - \|x\|^2)^2} .$$

(2) is proved in the same way, using the map F^{-1} instead of P. □

Note that both Riemannian metrics in Proposition 2.8 are conformal metrics: their expressions are a positive function times the Euclidean Riemannian metric of the underlying subset of $\mathbb{E}^n$.

The Riemannian structure defines a natural volume form and a volume measure on hyperbolic space: If V is for example an open subset of n-dimensional hyperbolic space, and λ_n is the n-dimensional Lebesgue measure, the volume of V is

$$\mathrm{Vol}(V) = \int_V \frac{2^n \, d\lambda_n(x)}{(1 - \|x\|^2)^n}$$

in the Poincaré ball model and

$$\mathrm{Vol}(V) = \int_V \frac{d\lambda_n(x)}{x_n^n}$$

in the upper halfspace model.

Example 2.9. As the isometry group acts transitively, the volume of each ball of a fixed radius is the same. Thus, it suffices to consider one ball that has a convenient center. The Euclidean radius of a ball of hyperbolic radius r centered at 0 in the Poincaré model is obtained by solving for R in the equation

$$r = d(0, (R, 0)) = \int_0^R \frac{2s}{1-s^2} = \log \frac{1+R}{1-R}.$$

This shows that the Euclidean radius of a hyperbolic ball centered at the origin of the Poincaré model is $\tanh \frac{r}{2}$. In order to compute the volume of the ball of radius r, recall that the Lebesgue measure is given in the spherical coordinates ($x \leftrightarrow (r, u)$) by $d\lambda_n(x) = r^{n-1} d\,\mathrm{Vol}_{\mathbb{S}^{n-1}}(u)$, and thus, using a change of variables $s \leftrightarrow \tanh \frac{t}{2}$, we get

$$\begin{aligned}
\mathrm{Vol}(\mathbb{B}(x,r)) = \mathrm{Vol}(\mathbb{B}(0,r)) &= \mathrm{Vol}(\mathbb{S}^{n-1}) \int_0^{\tanh \frac{r}{2}} \frac{2^n s^{n-1}}{(1+s^2)^n} ds \\
&= 2^{n-1}\,\mathrm{Vol}(\mathbb{S}^{n-1}) \int_0^r \sinh^{n-1} \frac{t}{2} \cosh^{n-1} \frac{t}{2}\, dt \\
&= \mathrm{Vol}(\mathbb{S}^{n-1}) \int_0^r \sinh^{n-1} t \, dt.
\end{aligned}$$

In the 2-dimensional case we easily get an explicit expression $\mathrm{Vol}(B^2(0,r)) = 4\pi \sinh^2 \frac{r}{2}$.

It is clear from the expression of the volume, that for all $x \in \mathbb{H}^n$, we have

$$\mathrm{Vol}(\mathbb{B}^n(x,r)) \sim \frac{\mathrm{Vol}(\mathbb{S}^n)}{2^{n-1}} e^{(n-1)r},$$

as $r \to \infty$. Thus, the volume of balls in hyperbolic space grows exponentially with the radius, much faster than in Euclidean space.

3. Geometry

In this section, we investigate a number of geometric properties of hyperbolic space using the various models according to the needs of the situation.

As the Riemannian metric of both the Poincaré ball model and the upper halfspace model are conformal with the Euclidean Riemannian structure, the angle between two tangent vectors is in these models is the same as the Euclidean angle. This allows us to prove the following facts on triangles in hyperbolic space. We say that a triangle is degenerate if it is contained in a geodesic segment.

Proposition 3.1. *(1) The sum of the angles of a nondegenerate triangle in hyperbolic space is strictly less than π.*
(2) For any $0 < \alpha, \beta, \gamma < \pi$ for which $\alpha + \beta + \gamma < \pi$, there is a triangle with angles α, β and γ.

Proof. Any three points in the hyperboloid model $\mathbb{H}^n$ are contained in the intersection of $\mathbb{H}^n$ with a 3-dimensional linear subspace of $\mathbb{R}^{1,n}$, which is an isometrically embedded copy of the hyperbolic plane. Furthermore, the geodesic arc between any two of these points in is contained in the same 2-plane. Thus, any triangle is always contained in an isometrically embedded copy of $\mathbb{H}^2$ in $\mathbb{H}^n$, so in the proof below, it suffices to consider the hyperbolic plane. We may assume that one of the vertices A

is the origin in the Poincaré disk model. Thus, two sides of the triangle are contained in two radii of the ball and the third one is contained in a circle which is orthogonal to the boundary of $\mathbb{B}(0,1)$.

(1) Consider the Euclidean triangle with the same vertices as T. The angles β and γ are strictly smaller than the corresponding angles in the Euclidean triangle. This implies the result as the angles of an Euclidean triangle sum to π.

(2) Sketch: Fix the side containing A and B to be contained in the positive real line. Then consider the family of circles C_s, $s \in [0,1[$ that are orthogonal to the Euclidean unit circle and form an angle β with the segment $[0,1[$ at the point of intersection s. When s is small, the side from B to C is very close (in the underlying Euclidean space) to the euclidean segment connecting B and C. When s increases, there is a unique parameter $0 < t < 1$ for which the circle C_t is tangent to the ray that forms an angle α with the positive real line. Continuity implies that all angles $0 < \gamma < \pi - \alpha - \beta$ are realised for some parameter in $]0, t[$. □

In the proof of the above result we made the following observation which is important in itself:

> If the sides of a triangle in hyperbolic space are all short, then the angle sum is almost π.

A related observation that uses the hyperbolic law of cosines, equality of angles and the second order Taylor polynomials of the hyperbolic functions is

> If the sides of a triangle in hyperbolic space are all short, then the sides satisfy the Euclidean law of cosines up to a small error.

3.1. **Polygons.** Any hyperplane P divides $\mathbb{H}^n$ into two open hyperbolic *halfspaces* which are the two components of the complement of P. If I is a finite or countable index set and $(H_i)_{i\in I}$ is a collection of closed halfplanes in $\mathbb{H}^n$ with nonempty intersection $P = \bigcap_{i\in I} H_i$ such that $(\partial H_i)_{i\in I}$ is a locally finite collection of hyperplanes (that is, for any compact $K \subset \mathbb{H}^n$, the set $\{i \in I : K \cap \partial H_i \neq \emptyset\}$ is finite), then P is a *locally finite polytope* in $\mathbb{H}^n$. In dimension $n = 2$, polytopes are called *polygons* and in dimension $n = 3$, *polyhedra*.

Three halfplanes in $\mathbb{H}^2$, whose pairwise intersections are nonempty and not halfplanes, define a polygon with three *sides* that are geodesic segments, rays or lines. On the other hand, any nondegenerate triangle in $\mathbb{H}^2$ defines a bounded polygon: The side $[A, B]$ is contained in a geodesic line L_C. Let H_C be the closed halfplane determided by L_C that contains C, and define the halfplanes H_A and H_B in the same way. Now, $P = H_A \cap H_B \cap H_C$ is a polygon, which we also refer to as a triangle. Note that there are more general polygons with three sides than the bounded ones:

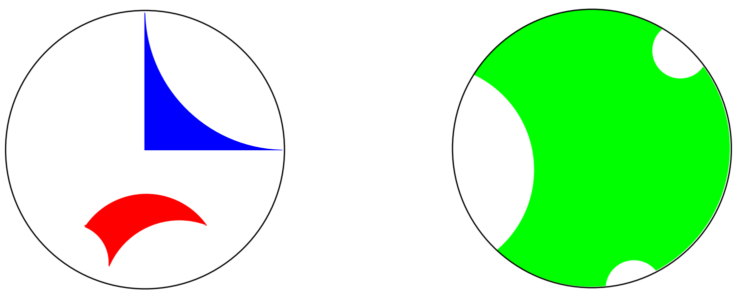

FIGURE 4

In Subsection 3.4, we will prove the following result:

Proposition 3.2. *The area of a triangle is* $\pi - (\alpha + \beta + \gamma)$.

Proposition 3.3. *Let* $n \geq 3$ *and let* $0 < \theta_1, \theta_2, \ldots, \theta_n < \pi$ *be angles that satisfy the equation*

$$\sum_{i=1}^{n} \theta_i < (n-2)\pi \,. \tag{6}$$

Then, there is a polygon with angles $\theta_1, \theta_2, \ldots, \theta_n$ *in this cyclic order around the boundary of the polygon.*

Proof. See [Bea, Thm. 7.16.1] for the complete proof. The proof for polygons, all of whose angles are equal, is more straightforward: Assume that $\theta_1 = \theta_2 = \cdots = \theta_n = \theta \in]0, \pi[$. The condition (6) implies that $\theta + 2\pi/n < \pi$, and thus, Proposition 3.1(2) implies that there is a unique triangle T with angles $\theta/2, \theta/2, 2\pi/n$. Let R be the rotation by angle $2\pi/n$ that fixes the origin in the disk model. This map is an isometry of the hyperbolic plane. If we place T such that the vertex with angle $2\pi/n$ is at the origin of the disc model and then the n copies $T, R(T), \ldots, R^{n-1}(T)$ of this triangle cover a neighbourhood of the origin and make up an n-gon, all of whose angles are equal to θ.

□

Note that unlike in Euclidean geometry, there are no right-angled quadrilaterals in $\mathbb{H}^2$. On the other hand, there are right-angled n-gons for all $n \geq 5$.

Proposition 3.4. *The area of a compact* n*-gon with angles* $\theta_1, \theta_2, \ldots, \theta_n$ *is*

$$(n-2)\pi - \sum_{i=1}^{n} \theta_i \,.$$

Proof. Let A be a vertex of an n-gon P. Divide P into triangles P_1, P_2, …, P_{n-2} with angles $\alpha_i, \beta_i, \gamma_i$ by adding the $n-3$ geodesic segments that connect P to the sides which are not adjacent to P in the boundary of P. The area of P_i is $\pi - (\alpha_i + \beta_i + \gamma_i)$. The angle sums of the triangles add up to $\sum_{i=1}^{n} \theta_i$, and their areas add up to the area of P. □

3.2. **Geodesic lines and isometries.** We already know that the geodesic lines of the upper halfspace model are, as sets, the intersections with $\mathbb{H}^n$ with Euclidean circles and lines that are orthogonal to $\mathbb{E}^{n-1} = \mathbb{E} \times \{0\}$. The following easy lemma records the expressions of the geodesics as mappings:

Lemma 3.5. *Let* $x \in \mathbb{R}^{n-1}$ *and* $y > 0$. *The maping* $\gamma \colon \mathbb{R} \to \mathbb{H}^n$,

$$\gamma(t) = (x, ye^t)$$

is a geodesic line in the upper halfspace model. For any isometry $g \in \operatorname{Isom} \mathbb{H}^n$, *the mapping* $g \circ \gamma$ *is a geodesic line.* □

In the upper halfspace model, it is often convenient to move a geodesic line by an isometry such that the endpoints of the geodesic in the model are 0 and ∞. The following results on isometries allow to do that and a bit more.

Lemma 3.6. *Let $a \in \mathbb{E}^{n-1} \times \{0\}$, and let $r > 0$.*

(1) *The inversion in the sphere $\mathbb{S}^{n-1}(a, r)$ preserves the upper halfspace and its restriction to the upper halfspace model is an isometry.*
(2) *The Euclidean reflection in a hyperplane orthogonal to $\mathbb{E}^{n-1} \times \{0\}$ preserves the upper halfplane and its restriction to the upper halfspace model is an isometry.*

Proof. Let us prove (1): The first claim is clear. To prove the second, it is enough to show that the expression $\frac{\|x-y\|^2}{x_n y_n}$ is invariant under the reflection. Now

$$f(x) - f(y) = \frac{x-a}{\|x-a\|^2} - \frac{y-a}{\|y-a\|^2} = \frac{(x-a)\|y-a\|^2 - (y-a)\|x-a\|^2}{\|x-a\|^2\|y-a\|^2},$$

which gives

$$\frac{\|f(x)-f(y)\|^2}{f(x)_n f(y)_n} = \frac{\|x-a\|^2\|y-a\|^4 - 2(x-a|y-a)\|x-a\|^2\|y-a\|^2 + \|x-a\|^4\|y-a\|^2}{\|x-a\|^4\|y-a\|^4} \frac{\|x-a\|^2\|y-a\|^2}{x_n y_n} = \frac{\|x-y\|^2}{x_n y_n}.$$

The proof of (2) is an exercise. □

Proposition 3.7. *Let x_1, x_2, x_3 and y_1, y_2, y_3 be two triples of distinct points in $\mathbb{R}^{n-1} \cup \{\infty\}$. There is an isometry of the upper halfspace model of $\mathbb{H}^n$ which is the restriction of a continuous map $g \colon \mathbb{R}^n \cup \{\infty\} \to \mathbb{R}^n \cup \{\infty\}$ such that $g(x_i) = y_i$ for all $i \in \{1, 2, 3\}$.*

Proof. The maps

- $T_b(x) = x + b$ for any $b \in \mathbb{R}^{n-1} = \mathbb{R}^n \times \{0\} \subset \mathbb{R}^{n+1}$, (horizontal translations)
- $\iota(x) = x/\|x\|^2$, (inversion in the Euclidean unit sphere)
- $L_\lambda(x) = \lambda x$ for any $\lambda > 0$ (dilation), and
- $Q(\bar{x}, x_n) = (Q_0(x), x_n)$ for any $Q \in \mathrm{O}(n-1)$

are isometries of the upper halfpace model by Proposition 3.6 and they are clearly continuous mappings of the one point compactification of $\mathbb{R}^n$ to itself.

It suffices to show that we can use the above isometries to map x_1, x_2, x_3 to $\infty, 0, (1, 0, \ldots, 0)$. The claim then follows by composing such a map with the inverse of another one. If all points x_1, x_2, x_3 are finite, map x_1 by a translation T_{-x_1} to 0 and then by the inversion ι to ∞. Rename $\iota \circ T_{-x_1}(x_2)$ and $\iota \circ T_{-x_1}(x_3)$ to x_2 and x_3. Map x_2 to 0 by a translation. This map keeps ∞ fixed. Map x_3 (again renamed) to the unit sphere by a dilation and then to $(1, 0, \ldots, 0)$ by the extension of an orthogonal map of $\mathbb{E}^{n-1}$. These two maps fix ∞ and 0. □

Corollary 3.8. *Let x_1, x_2, x_3 and y_1, y_2, y_3 be two triples of distinct points in $\mathbb{S}^{n-1} \cup \{\infty\}$. There is an isometry of the ball model of $\mathbb{H}^n$ which is the restriction of a continuous map $g \colon \mathbb{E}^n \cup \{\infty\} \to \mathbb{E}^n \cup \{\infty\}$ such that $g(x_i) = y_i$ for all $i \in \{1, 2, 3\}$.* □

The proof of Proposition 3.7 gives the following useful corollary:

Corollary 3.9. *(1) The subgroup of* $\mathrm{Isom}(\mathbb{H}^n)$ *generated by dilations and translations acts transitively on $\mathbb{H}^n$.*

(2) The subgroup of $\mathrm{Isom}(\mathbb{H}^n)$ *generated by dilations and orthogonal maps acts transitively on $\mathbb{H}^n$.* □

Lemma 3.10. *Let L be a geodesic line in $\mathbb{H}^n$, and let $p \in \mathbb{H}^n$. There is a unique point $\pi_L(p) \in L$ for which $d(p, L) = d(p, \pi_L(p))$. Furthermore, if $\gamma \colon \mathbb{R} \to \mathbb{H}^n$ is a geodesic line, then the function $s \mapsto d(\gamma(s), p)$ is convex.*

Proof. Using Proposition 3.7, we may assume that L is the geodesic line with endpoints 0 and ∞ in the upper halfspace model. There is an isometrically embedded copy of $\mathbb{H}^2$ in $\mathbb{H}^n$ that contains L, p and any geodesic line through p and an point of L. Thus, we can assume that $n = 2$. Using a dilation, we can assume that $\|p\| = 1$, which means that $p = (\cos\theta, \sin\theta)$ for some $0 < \theta < \pi$. For any $s > 0$, we have

$$\cosh d(e^s, p) = 1 + \frac{\cos^2\theta + (e^s - \sin\theta)^2}{2e^s \sin\theta} = \frac{\cosh s}{\sin\theta} \geq \frac{1}{\sin\theta} = \cosh((0,1), p)\,.$$

The function $s \mapsto \operatorname{arcosh}(\cosh(s)/\sin\theta)$ is easily seen to be convex. □

The map $\pi_L \colon \mathbb{H}^n \to L$ is the *closest point map* of L. Note that for any $p \notin L$, the geodesic arc $[p, \pi_L(p)]$ is orthogonal to L.

Proposition 3.11. *The closest point map is* 1*-Lischitz. More precisely, for any* $p, q \in \mathbb{H}^n$,

$$d(\pi_L(p), \pi_L(q)) \leq d(p, q)$$

with equality only if $p, q \in L$.

Proof. Normalize so that L is the geodesic line with endpoints at 0 and ∞. For any $x \in L$,

$$\pi_L^{-1}(x) = \{y \in \mathbb{H}^n : \|y\| = \|x\|\}\,.$$

Let us minimize the hyperbolic distance between any pair of points $x, y \in \mathbb{H}^n$ such that $\pi_L(x) = \pi_L(p)$ and $\pi_L(y) = \pi_L(q)$. The Euclidean distance $\|x - y\|$ is minimal when x and y are on the same ray from 0 and the product $x_n y_n$ is maximal when x and y are on L. The claim now follows from the expression (5) of the hyperbolic metric. □

For any $r > 0$, the *r-neighbourhood* of any nonempty subset $A \subset \mathbb{H}^n$ is

$$\mathcal{N}_r(A) = \{x \in \mathbb{H}^n : d(x, A) < r\}\,.$$

The following obsevation is a straightforward exercise:

Lemma 3.12. *The* r*-neighbourhood of the geodesic line* L *with endpoints* 0 *and* ∞ *in the upper halfspace model is the Euclidean cone*

$$\mathcal{N}_r(L) = \{x \in \mathbb{H}^n : \cos\angle_0(L, x) > \frac{1}{\cosh r}\}\,. \quad \square$$

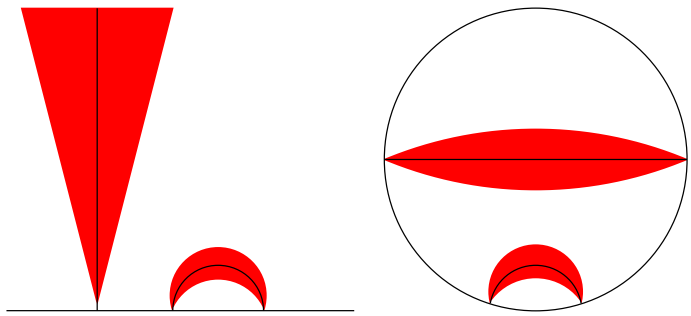

FIGURE 5

If L' is a geodesic line in the upper halfspace model, we can map it to L by a composition of the isometries used in Proposition 3.7. These isometries are conformal maps which map the set of spheres and hyperplanes in $\mathbb{R}^n \cup \{\infty\}$ to itself. It is easy to see that the neighbourhoods $\mathcal{N}_r(L')$ are cones or bananas with opening angles at the endpoints given by Lemma 3.12, see Figure 5. As the isometry used to map the ball model to the upper halfspace model is an inversion, the r-neighbourhoods of geodesic lines in the ball model are bananas.

3.3. **The boundary at infinity.** In the Poincaré ball model, hyperbolic space appears to have a boundary, the Euclidean unit sphere. However, this set is not a subset of hyperbolic space, and it is not difficult to see that $\mathbb{H}^n$ is a complete metric space (exercise, use the fact that closed balls are compact). However, the unit sphere $\mathbb{S}^{n-1}$ has a natural geometric meaning as the boundary of infinity of $\mathbb{H}^n$, as we will now see:

Two geodesic rays $\rho_1, \rho_2 \colon \mathbb{R} \to \mathbb{H}^n$ are *asymptotic*, if

$$\sup_{t\in[0,\infty]} d(\rho_1(t), \rho_2(t)) < \infty\,.$$

Asymptoticity defines an equivalence relation $\sim$ in the set of geodesic rays of $\mathbb{H}^n$. The set

$$\partial_\infty \mathbb{H}^n = \{\text{geodesic rays } \rho\colon \mathbb{R} \to \mathbb{H}^n\}/\sim$$

is the *boundary at infinity* of $\mathbb{H}^n$.

Proposition 3.13. *Two geodesic rays $\rho_1, \rho_2 \colon \mathbb{R} \to \mathbb{H}^n$ are asymptotic if and only if they have the same endpoint in the Poincaré ball model or in the upper halfspace model.*

Proof. As the inversion used to identify the two models in Subsection 2.3 is a continuous map of $\mathbb{R}^n \cup \{\infty\}$, it suffices to consider the upper halfspace model.

Assume that the geodesic rays ρ_1 and ρ_2 have the same endpoint in the upper halfspace model. Using Proposition 3.7, we can assume the endpoint is ∞. Now, there are $\bar{x}, \bar{y} \in \mathbb{E}^{n-1}$ and $x_n, y_n > 0$ such that $\rho_1(t) = (\bar{x}, x_n e^t)$ and $\rho_2(t) = (\bar{y}, y_n e^t)$. We estimate the distance in the definition of asymptoticity:

$$d(\rho_1(t), \rho_2(t)) \le |\log \frac{x_n e^t}{y_n e^t}| + \frac{\|\bar{x} - \bar{y}\|}{\max(x_n, y_n)e^t} \le |\log \frac{x_n}{y_n}| + \frac{\|\bar{x} - \bar{y}\|}{\max(x_n, y_n)}|\,,$$

which implies asymptoticity.

If the rays ρ_1 and ρ_2 have different endpoints in the model, we can assume that these points are ∞ and 0. Now, ρ_1 is as above and $\rho_2(t) = (0, y_n e^{-t})$. Clearly,

$$d(\rho_1(t), \rho_2(t)) \ge d(\rho_1(t), (0, x_n e^t))$$

for all $t \in \mathbb{R}_+$. Thus, we have for big t,

$$d(\rho_1(t), \rho_2(t)) \ge 2t - |\log x_n - \log y_n| \to \infty$$

as $t \to \infty$. This shows that the rays are not asymptotic. □

Corollary 3.14. *The boundary at infinity of $\mathbb{H}^n$ is naturally identified with $\mathbb{S}^{n-1}$.*

Proposition 3.15. *Any isometry of $\mathbb{H}^n$ extends to a bijection $\bar{g}$ of $\partial_\infty \mathbb{H}^n$ defined by*

$$\bar{g}([\rho]) = [g \circ \rho]\,.$$

Proof. If ρ_1 and ρ_2 are geodesic rays and $g \in \operatorname{Isom} \mathbb{H}^n$ is an isometry, then for all $t \in \mathbb{R}$, we have

$$d(g \circ \rho_1(t), g \circ \rho_2(t)) = d(\rho_1(t), \rho_2(t))\,.$$

Thus the boundary map is well-defined. The fact that it is a bijection follows from similar computations. □

We can now express Proposition 3.7 in a more intrinsic way:

Proposition 3.16. *Let x_1, x_2, x_3 and y_1, y_2, y_3 be two triples of distinct points in $\partial_\infty \mathbb{H}^n$. There is an isometry g of $\mathbb{H}^n$ such that $\bar{g}(x_i) = y_i$ for all $i \in \{1, 2, 3\}$.* □

3.4. **Triangles with vertices at infinity.** We now extend the definition of triangles and allow some of the vertices to be points at infinity of $\mathbb{H}^n$: A *(generalized) triangle* consists of three distinct points $A, B, C \in \mathbb{H}^n \cup \partial_\infty \mathbb{H}^n$, called the *vertices*, and of the geodesic arcs, rays or lines, called the *sides*, connecting the vertices.

We begin by proving an analog of the second law of cosines for a special kind of generalized triangles. Note that the first law of cosines cannot be generalized to this setting as the triangle in question has two infinitely long sides.

Proposition 3.17. *Let $A, B \in \mathbb{H}^n$ and let $C \in \partial_\infty \mathbb{H}^n$. Then*

$$\cosh c = \frac{1 + \cos\alpha \cos\beta}{\sin\alpha \sin\beta}. \tag{7}$$

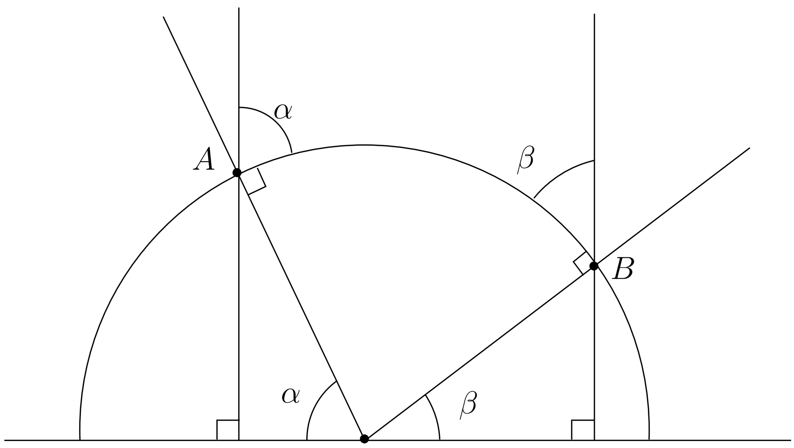

FIGURE 6

Proof. We use the upper halfplane model and normalise so that A and B are on the Euclidean unit circle and $C = \infty$. Assume that $\arg B < \frac{\pi}{2} \leq \arg A$. Then $A = (-\cos\alpha, \sin\alpha)$ and $B = (\cos\beta, \sin\beta)$. The result follows from equation (5). ☐

The special case of equation (7) with $\beta = \frac{\pi}{2}$:

$$\cosh c = \frac{1}{\sin\alpha}$$

is known as the *angle of parallelism.*

Note that equation (7) agrees with the second law of cosines if we define that

the angle at a vertex at infinity is 0.

From now on, we will use this convention.

Proposition 3.2 is contained in the following, more general result:

Proposition 3.18. *The area of a (generalized) triangle is $\pi - (\alpha + \beta + \gamma)$.*

Proof. Any triangle T can be described as the difference of two triangles with at least one vertex at infinity. By the additivity of area and angles in the hyperbolic plane, we may restrict to this special case. Using Proposition 3.7, we can assume that $C = \infty$ and that A and B are on the Euclidean unit circle, possibly in the boundary at infinity. Now, the area of T is

$$\int_T \frac{d\lambda_2(x)}{x_2^2} = \int_{-\cos(\alpha)}^{\cos\beta} \int_{\sqrt{1-x_1^2}}^{\infty} \frac{dx_1 dx_2}{x_2^2} = \int_{\cos(\pi-\alpha)}^{\cos\beta} \frac{dx_1}{\sqrt{1-x_1^2}} = \pi - \alpha - \beta\,. \quad \square$$

A triangle, all of whose angles are in the boundary at infinity, is called an *ideal triangle.* Proposition 3.18 implies that the area of an ideal triangle is π. Note also that the area of any triangle is at most π.

3.5. **Balls and horoballs.** In this section, we observe that hyperbolic balls in the upper halfspace model and in the Poincaré ball model are balls in the Euclidean metric. Furthermore, we give a gemetric interpretation to the Euclidean balls in these models that are tangent to the boundary at infinity.

Proposition 3.19. *Balls in the upper halfspace model and in the Poincaré ball model are Euclidean balls in the model space.*

Proof. By symmetry, balls centered at the origin of the ball model are Euclidean balls. The inversion that maps the ball model to the upper halfspace model is an isometry, and on the other hand it preserves generalized spheres. Thus, the images of the balls centered at the origin are hyperbolic and Euclidean balls. The hyperbolic center of these balls can be mapped to any other point in $\mathbb{H}^n$ by one of the isometries used in the proof of Proposition 3.7. These mappings preserve generalized spheres, which implies that all balls in the upper halfspace model are Euclidean balls. The rest of the claim follows by one more application of the inversion that maps the ball model to the upper halfspace model. $\square$

Note that the Euclidean radii and centers of the balls hardly ever coincide with the hyperbolic ones.

Next, we will give a geometric meaning to the Euclidean balls and halfspaces that are contained in the ball or upper halfspace models and are tangent to the sphere at infinity. Note that these subsets cannot be balls because their diameters are infinite.

Let $p \in \mathbb{H}^n$ and $\xi \in \partial_\infty \mathbb{H}^n$. The *Busemann function* of ξ with respect to p is the function $\beta_{\xi,p} \colon \mathbb{H}^n \to \mathbb{R}$,

$$\beta_{\xi,p}(q) = \lim_{t\to\infty} (t - d(\rho(t), q)) = \lim_{t\to\infty} (d(\rho(t), p) - d(\rho(t), q)),$$

where ρ is the geodesic ray whose endpoint at infinity is ξ and such that $\rho(0) = p$. The superlevel sets $\beta_{\xi,p}^{-1}([t, \infty[)$ are (closed) *horoballs*, and the level sets $\beta_{\xi,p}^{-1}(t)$ are *horospheres* centered at ξ.

Proposition 3.20. *(1) The Busemann function is well-defined.*
(2) In the upper halfspace model, horoballs are Euclidean balls tangent to $\mathbb{R}^{n-1}$ *or Euclidean halfspaces* $\{x \in \mathbb{H}^n : x_n \geq c\}$*. In the ball model, horoballs are Euclidean balls tangent to* $\mathbb{S}^{n-1}$*.*

Proof. The Busemann function is defined in terms of the hyperbolic metric, and thus, it suffices to consider the case $\xi = \infty$ and $p = (0, 1) \in \mathbb{R}^{n-1} \times \mathbb{R}_+ = \mathbb{H}^n$. Let $\rho(t) = (\bar{p}, p_n e^t)$ and let $x = (\bar{x}, x_n)$. A computation using, for example, the expression (5) of the hyperbolic metric shows that $t - d(\rho(t), x)$ converges to $\log p_n - \log x_n$ as $t \to \infty$. This proves the claims. $\square$

Let L be a geodesic line and let $\xi \in \partial_\infty \mathbb{H}^n$ be one of its endpoints. The *horocyclic projection map* of L with respect to ξ is the mapping $h_{L,\xi} \colon \mathbb{H}^n \to L$ defined by setting $h_{L,\xi}(p)$ to be the unique point on L which is on the horocycle centered at ξ that passes through p. In other words, $\beta_{\xi,p}(h_L(p)) = 0$. Note that if η is the other endpoint of L, the maps $h_{L,\xi}$ and $h_{L,\eta}$ are not the same.

Proposition 3.21. *The horocyclic projection is* 1*-Lipschitz. More precisely, for any* $p, q \in \mathbb{H}^n$,

$$d(\pi_L(p), \pi_L(q)) \leq d(p,q)$$

with equality only if $p, q \in L$.

Proof. Exercise. □

In the proof of the second claim of Proposition 3.13 we employed the horocyclic projection to the geodesic line that contains one of the geodesic rays studied in the proof, with respect to the endpoint ∞.

4. Möbius transformations and isometries

4.1. **The Poincaré extension.** In this section, we take a closer look at the isometries of the upper halfspace model and the Poincaré model. Let $\widehat{\mathbb{E}}^n = \mathbb{E}^n \cup \{\infty\}$ be the one point compactification of $\mathbb{E}^n$. The set of generalized hyperplanes (or generalized spheres) of $\widehat{\mathbb{E}}^n$ consists of the Euclidean spheres and the hyperplanes with the point ∞ added to each hyperplane. This is a reasonable convention because the image of a generalized sphere under an inversion is a generalised sphere, and the same holds for Euclidean reflections in hyperplanes, see Appendix **??**. In this context, both inversions in spheres and Euclidean reflections in hyperplanes are referred to as *reflections (in generalized hyperplanes)* or as *inversions (in generalized spheres)*.

The group $\text{Möb}(\widehat{\mathbb{E}}^n)$ generated by all reflections in generalized hyperplanes of $\widehat{\mathbb{E}}^n$ is the *general Möbius group* of $\widehat{\mathbb{E}}^n$. Its elements are called *Möbius transformations.* The *general Möbius group* $\text{Möb}(\mathbb{S}^{n-1})$ of $\mathbb{S}^{n-1}$ is the subgroup of $\text{Möb}(\widehat{\mathbb{E}}^n)$ generated by the restrictions to $\mathbb{S}^{n-1}$ of the reflections in the spheres in $\widehat{\mathbb{E}}^n$ that are orthogonal to $\mathbb{S}^{n-1}$.

Example 4.1. (1) Translations $x \mapsto x + a$, $a \in \mathbb{E}^n$, are compositions of reflections in two parallel hyperplanes: If

$$R_H = R_{x_0,u}(x) = x - 2(x - x_0|u)u$$

is the reflection in the hyperplane $H = x_0 + u^\perp$, then

$$R_{x_2,u} \circ R_{x_1,u}(x) = x + 2(x_2 - x_1|u)u$$

is a translation, and we can choose x_1, x_2 and u such that $2(x_2 - x_1|u)u = a$.
(2) Orthogonal maps are are compositions of reflections in hyperplanes that intersect at the origin.
(3) Dilations $x \mapsto \lambda x$, $\lambda > 0$ are compositions of inversions in two concentric spheres: For any $\alpha, \beta > 0$, we have

$$\iota_{0,\beta} \circ \iota_{0,\alpha} = \frac{\beta}{\alpha} x \,.$$

The results in Appendix **??** imply the following basic properties:

Proposition 4.2. *Möbius transformations are conformal. They map generalized hyperplanes to generalized hyperplanes.*

The following result justifies considering spheres and hyperplanes as one family:

Theorem 4.3. *If H is a generalized sphere and ϕ is a Möbius transformation that fixes H pointwise. Then ϕ is the reflection in H or the identity map. Furthermore, any two reflections are conjugate in the general Möbius group.*

Proof. See [Bea, Thm. 3.2.4]. □

In particular, Theorem 4.3 shows that inversions in spheres and reflection in hyperplanes are completely analogous transformations in $\widehat{\mathbb{E}}^n$: Let $x \in \mathbb{S}(a, R) \subset \mathbb{E}^n$. The inversion $\iota_{x,r}$, for any $r > 0$, maps $\mathbb{S}(a, R)$ to a hyperplane $H \cup \{\infty\}$. The map $\iota_{x,r} \circ \iota_{a,R} \circ \iota_{x,r}$ is a Möbius transformation which is not the identity map and which fixes $H \cup \{\infty\}$ pointwise. Theorem 4.3 implies that $\iota_{x,r} \circ \iota_{a,R} \circ \iota_{x,r} = R_H$.

Lemma 3.6 implies that we can extend any Möbius transformation of $\widehat{\mathbb{E}}^{n-1}$ to an isometry of the upper halfplane model of $\mathbb{H}^n$ as follows: A Möbius transformation $A \in \text{Möb}(n-1)$ is the composition of reflections $r_1, \dots, r_k$ in generalized spheres of $\widehat{\mathbb{E}}^{n-1}$: $A = r_1 \circ \cdots \circ r_k$. Let $\widetilde{r}_1, \dots \widetilde{r}_k$ be the reflections in the generalized spheres of $\widehat{\mathbb{E}}^n$ that intersect $\widehat{\mathbb{E}}^{n-1}$ perpendicularly at the fixed generalized spheres of the reflections $\widetilde{r}_1, \dots \widetilde{r}_k$. The *Poincaré extension* of the Möbius transformation A is the isometry $\widetilde{A} = \widetilde{r}_1 \circ \cdots \circ \widetilde{r}_k$ of the upper halfspace model. Similarly, any Möbius transformation $B \in \text{Möb}(\mathbb{S}^{n-1})$ is the composition of the restrictions to $\mathbb{S}^{n-1}$ of the reflections in spheres $S_1, \dots, S_m$ orthogonal to $\mathbb{S}^{n-1}$, and the *Poincaré extension* of M, defined as the composition of the reflections in the spheres $S_1, \dots, S_m$ is seen to be an isometry of the ball model, using Lemma 3.6 and the inversion that maps the ball model to the upper halfspace model.

Proposition 4.4. *The group Möb$(n-1)$ acts transitively by isometries on $\mathbb{H}^n$.*

Proof. Use Corollary 3.9 and Example 4.1. □

Theorem 4.5. *(1) The isometry group of the n-dimensional Poincaré model is the Poincaré extension of Möb$(\mathbb{S}^{n-1})$.*
(2) The isometry group of the n-dimensional Poincaré model is the Poincaré extension of Möb$(\mathbb{E}^{n-1} \cup \{\infty\})$.

Proof. (1) We will use the map P from the ball model to the hyperboloid model defined in section 2.2. Recall that we know the isometry group of the hyperboloid model and the stabilizer of each point in it from Section 1.

The generalised $(n-1)$-spheres that are orthogonal to $\mathbb{S}^{n-1}$ and pass through the origin of the ball model and are orthogonal to the unit sphere are exactly the linear $(n-1)$-planes of $\mathbb{E}^n$. The Euclidean reflections in these hyperplanes are isometries of the Poincaré model. Thus, the stabilizer of 0 in the isometry group contains a copy of $\mathrm{O}(n)$. It is straightforward to check that if Q is any orthogonal transformation of $\mathbb{E}^n$, then P conjugates Q to $\text{diag}(1, Q) \in \text{Stab}_{\text{Isom}(\mathbb{H}^n)}(1, 0)$. Since $\text{Isom}(\mathbb{H}^n)$ and $\text{Möb}(n-1)$ act transitively by isometries on, respectively, the hyperboloid model and the ball model, the claim follows.
(2) The map S conjugates the isometries of the ball model with those of the upper halfspace model, and it conjugates the groups of Möbius transformations $\text{Möb}(\mathbb{S}^{n-1})$ and $\text{Möb}(\widehat{\mathbb{E}}^{n-1})$ with each other. □

4.2. **Isometry groups in low dimensions.** We can use complex Möbius transformations to describe isometries of the hyperbolic plane and hyperbolic 3-space.

The *special linear groups* with real and complex coefficients are

$$\text{SL}_2(\mathbb{R}) = \{A \in \text{GL}_2(\mathbb{R}) : \det A = 1\}$$

and

$$\mathrm{SL}_2(\mathbb{C}) = \{A \in \mathrm{GL}_2(\mathbb{C}) : \det A = 1\},$$

and the *special unitary group* of signature $(1,1)$ is

$$\mathrm{SU}(1,1) = \{A \in \mathrm{SL}_2(\mathbb{C}) : A^* J A = J\} = \left\{ \begin{pmatrix} a & b \\ \bar{b} & \bar{a} \end{pmatrix} : |a|^2 - |b|^2 = 1 \right\},$$

where $J = \mathrm{diag}(-1.1)$ and $A^* = {}^T\bar{A}$.

Recall from complex analysis (see for example [Ahl]) that any matrix $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ with $a, b, c, d \in \mathbb{C}$ and $\det A \neq 0$ determines a Möbius transformation $A \colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$,

$$Az = \frac{az + b}{cz + d}. \tag{8}$$

The kernel of this action is $\{\pm \mathrm{id}\}$. It is easy to check that the mappings defined by equation 8 are indeed Möbius transformations in the sense of the definition in section 4.1. This follows, for example from the following result and Example 4.1, which shows that the mappings

$$z \mapsto -\frac{1}{z} = -\overline{\left(\frac{1}{\bar{z}}\right)}$$

and

$$z \mapsto z + b$$

that appear in Proposition 4.6 are Möbius transformations.

Proposition 4.6. *Let $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$. The group $\mathrm{SL}_2(\mathbb{K})$ is generated by the elements*

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 1 & \beta \\ 0 & 1 \end{pmatrix},$$

with $\beta \in \mathbb{K}$.

Proof. Assume $\alpha, \beta, \gamma, \delta \in \mathbb{K}$ with $\gamma \neq 0$. Then, since $\alpha\delta - \beta\gamma = 1$, we have $-\beta + \alpha\delta\gamma^{-1} = \gamma^{-1}$, and therefore, we have the following equation in $\mathrm{SL}_2(\mathbb{K})$:

$$\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} = \begin{pmatrix} 1 & \alpha\gamma^{-1} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \gamma & 0 \\ 0 & -\beta + \alpha\delta\gamma^{-1} \end{pmatrix} \begin{pmatrix} 1 & \gamma^{-1}\delta \\ 0 & 1 \end{pmatrix}.$$

The claim now follows from observing that for any $\alpha \in \mathbb{K} - \{0\}$, $\beta \in \mathbb{K}$, we have

$$\begin{pmatrix} \alpha & \beta \\ 0 & \alpha^{-1} \end{pmatrix} = \begin{pmatrix} \alpha & 0 \\ 0 & \alpha^{-1} \end{pmatrix} \begin{pmatrix} 1 & \beta\alpha^{-1} \\ 0 & 1 \end{pmatrix}$$

and

$$\begin{pmatrix} \alpha & 0 \\ 0 & \alpha^{-1} \end{pmatrix} = \begin{pmatrix} 1 & -\alpha \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & -\alpha^{-1} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & -\alpha \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}. \qquad \square$$

Let $f \in \mathrm{M\ddot{o}b}(\widehat{\mathbb{E}}^1)$, $f(x) = \frac{ax+b}{cx+d}$ be a Möbius transformation. The Poincaré extension of f to $\mathbb{H}^2$ is $f \colon \mathbb{H}^2 \to \mathbb{H}^2$, $f(z) = \frac{az+b}{cz+d}$. This can be seen, for example, using the Proposition 4.6 and observing that the reflections in the Euclidean unit circle and the imaginary axis have simple expressions using complex numbers

$$\iota_{0,1}(z) = \frac{z}{|z|^2} = \frac{1}{\bar{z}}$$

and

$$R_{i\mathbb{R}}(z) = z - 2\,\mathrm{Re}\, z = -\bar{z}\,.$$

Similarly, the Poincaré extension of any element of $\mathrm{SU}(1,1)$ to the ball model is given by equation (8).

There is an explicit expression for the Poincaré extension of elements of $\mathrm{SL}_2(\mathbb{C})$ to isometries of the upper halfspace model of $\mathbb{H}^3$ using Hamiltonian quaternions: If we write an element of the upper halfspace as (z, tj), $t > 0$ and j is a quaternion that satisfies $j^2 = -1$, $ij = -ji$, then the Poincaré extension is given by

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} (z, tj) = \frac{(az+b)\overline{(cz+d)} + a\bar{c}t^2 + tj}{|cz+d|^2 + |c|^2 t^2} .$$

See for example [Bea, Thm. 4.1.1]

Proposition 4.7. *(1)* $\mathrm{PSL}_2(\mathbb{R}) = \mathrm{SL}_2(\mathbb{R})/\{\pm \mathrm{id}\}$ *is the subgroup of index* 2 *in the isometry group of the upper halfplane model of* $\mathbb{H}^2$ *that consists of the orientation-preserving isometries.*

(2) $\mathrm{PSL}_2(\mathbb{C}) = \mathrm{SL}_2(\mathbb{C})/\{\pm \mathrm{id}\}$ *is the subgroup of index* 2 *in the isometry group of the upper halfspace model of* $\mathbb{H}^3$ *that consists of the orientation-preserving isometries.*

(3) $\mathrm{PU}(1,1) = \mathrm{SU}(1,1)/\{\pm \mathrm{id}\}$ *is the subgroup of index* 2 *in the isometry group of the Poincaré disk model of* $\mathbb{H}^2$ *that consists of the orientation-preserving isometries.*

Proof. (1) The isometries that are defined by elements of $\mathrm{SL}_2(\mathbb{R})$ are analytic functions, thus they are orientation-preserving. For any $A \in \mathrm{SL}_2(\mathbb{R})$, the mapping $\bar{A} = A \circ R_{i\mathbb{R}} = R_{i\mathbb{R}} \circ A \colon \mathbb{H}^2 \to \mathbb{H}^2$,

$$\bar{A}(z) = \frac{-a\bar{z} + b}{-c\bar{z} + d} \tag{9}$$

is an orientation-reversing isometry of the upper halfplane. Furthermore, if f is an isometry of $\mathbb{H}^2$, then $f = A$ or $f = \bar{A}$ for some $A \in \mathrm{SL}_2(\mathbb{R})$ because any reflection in a generalized sphere orthogonal to $\widehat{\mathbb{E}}^1$ is of the form and it is easy to check that for any $A, B \in \mathrm{SL}_2(\mathbb{R})$, $\bar{A} \circ \bar{B} = AB \in \mathrm{SL}_2(\mathbb{R})$.

Let $\Phi\colon \mathrm{Isom}(\mathbb{H}^2) \to \mathbb{Z}/2\mathbb{Z}$ be the mapping which associates $k \mod 2$ to $f \in \mathrm{Isom}(\mathbb{H}^2)$ if f can be written as a composition of reflections $f = r_1 \circ f_2 \circ \cdots \circ r_k$. One can check that Φ is well defined, and clearly, Φ is then a surjective homomorphism. The kernel of Φ is $\mathrm{PSL}_2(\mathbb{R}) < \mathrm{Isom}(\mathbb{H}^2)$, which proves $[\mathrm{Isom}(\mathbb{H}^2) : \mathrm{PSL}_2(\mathbb{R})] = 2$.

(2) See [Bea, Sect. 4.1].

(3) One can check by a direct calculation using the epression of the hyperbolic metric of the ball model given in section 2.2 that $\mathrm{PU}(1,1) < \mathrm{Isom}(\mathbb{H}^2)$. The proof is similar to that of (1). □

Note that, in fact, we found all orientation-preserving isometries of $\mathbb{H}^2$ and $\mathbb{H}^3$ in our proof of Proposition 3.7.

Remarks 4.8. (1) The trace of a matrix is invariant under conjugation:

$$\mathrm{tr}(BAB^{-1}) = \mathrm{tr}\, A$$

for all $A, B \in \mathrm{PSL}_2(\mathbb{C})$. Since the kernel of the map from $\mathrm{PSL}_2(\mathbb{C})$ to $\mathrm{Isom}(\mathbb{H}^n)$ is $\pm I_2$, the traces of the two matrices associates with an orientarion-preserving isometry differ by a sign, we can define a map $\mathrm{tr}^2\colon \mathrm{Isom}_+(\mathbb{H}^3) \to \mathbb{R}_+$. This map is invariant under conjugation, and it classifies the elements of $\mathrm{PSL}_2(\mathbb{R})$ and $\mathrm{PSL}_2(\mathbb{C})$ in three types. Items (2) to (4) below elaborate on the classification of $\mathrm{PSL}_2(\mathbb{R})$.

(2) Using the representation of orientation-preserving isometries of $\mathbb{H}^2$ by Möbius transformations, it is straightforward to check that an orientation-preserving isometry A of $\mathbb{H}^2$ which is not the identity has one or two fixed points in $\mathbb{H}^2 \cup \partial_\infty \mathbb{H}^2$.

More precisely, the fixed points of the transformation $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ are

$$\frac{a-d}{2c} \pm \frac{\sqrt{\operatorname{tr}^2 A - 4}}{2c}\,.$$

From this formula, we see that an isometry $A \in \mathrm{PSL}_2(\mathbb{R})$ has

- no fixed points in $\mathbb{H}^2$ and two fixed points in $\partial_\infty \mathbb{H}^2$ if $\operatorname{tr}^2 A > 4$,
- one fixed point in $\mathbb{H}^2$ and no fixed points in $\partial_\infty \mathbb{H}^2$ if $\operatorname{tr}^2 A \in [0, 4[$, and
- no fixed points in $\mathbb{H}^2$ and one fixed point in $\partial_\infty \mathbb{H}^2$ if $\operatorname{tr}^2 A > 4$.

(3) Using the above results, one can show that any two Möbius transformations $A, B \in \mathrm{SL}_2(\mathbb{R})$ or $A, B \in \mathrm{SL}_2(\mathbb{C})$ with $\operatorname{tr}^2 A = \operatorname{tr}^2 B$ are conjugate in $\mathrm{Isom}(\mathbb{H}^2)$ and $\mathrm{Isom}(\mathbb{H}^3)$, respectively. For example, if $A \in \mathrm{SL}_2(\mathbb{R})$ with $\operatorname{tr}^2 A > 4$, then A has two fixed points, which we may assume are 0 and ∞. Now, the equations for fixed points and the determinant imply that $A = \operatorname{diag}(\lambda, \lambda^{-1})$, which implies that $\operatorname{tr}^2 A = (\lambda + 1/\lambda)^2$. Conjugating with the map $z \mapsto -1/z$, we may assume that $\lambda > 1$. Similarly, B is conjugate with $\operatorname{diag}(\lambda, \lambda^{-1})$. The other cases are proved in a similar way.

(4) The above remark implies that An isometry $A \in \mathrm{PSL}_2(\mathbb{R}) - \{\mathrm{id}\}$ stabilizes

- the geodesic line that connects its two endpoints if $\operatorname{tr}^2 A > 4$,
- all hyperbolic balls centered at its fixed point if $\operatorname{tr}^2 A \in [0, 4[$, and
- all horoballs based at its fixed point if $\operatorname{tr}^2 A > 4$.

In section 1.3, we introduced three types of isometries, elliptic, hyperbolic and parabolic. Orientation-preserving isometries can be classified into these three categories according to the trace and, by the above observation, according to the number and nature of their fixed points: An isometry $A \in \mathrm{PSL}_2(\mathbb{R}) - \{\mathrm{id}\}$ is

- hyperbolic if $\operatorname{tr}^2 A > 4$,
- elliptic if $\operatorname{tr}^2 A \in [0, 4[$, and
- parabolic if $\operatorname{tr}^2 A > 4$.

The first two cases are easier, for the claim on parabolic elements, one can compute an expression of the Busemann function in the hyperboloid model to see that the set we called a horosphere in section 1.3 is indeed a horosphere according to the definition given in section 3.5.

5. Hyperbolic manifolds and discontinuous groups

In this section, we study quotient spaces of hyperbolic space under the action of discontinuous groups of isometries. If the elements of the groups in question have no fixed points in hyperbolic space, then the quotient spaces are metric spaces which are locally isometric with hyperbolic space.

A group $G < \mathrm{Isom}\,\mathbb{H}^n$ acts *discontinuously* on $\mathbb{H}^n$, if for any compact subset $K \subset \mathbb{H}^n$, the set $\{g \in G : g(K) \cap K \neq \emptyset\}$ is finite. If G acts discontinuously, we say that G is a *discontinuous group* of isometries.

Example 5.1. The following cyclic subgroups of $\mathrm{Isom}(\mathbb{H}^n)$ are discontinuous. Note that in the case of the hyperbolic plane, the examples corresponds to the three types that we considered in Examples 4.8.

(1) Let $\lambda > 0$, $\lambda \neq 1$, and let $L_\lambda(x) = \lambda x$. The group

$$\langle L_\lambda \rangle = \{L_\lambda^k x : k \in \mathbb{Z}\}$$

is a discontinuous group of isometries of the upper halfspace model of $\mathbb{H}^n$. The half-annuli

$$A_k = \{x \in \mathbb{H}^n : \lambda^k \leq \|x\| < \lambda^{k+1}\} = L_\lambda^k A_0$$

are disjoint and they contain exactly one point from each $\langle L_\lambda \rangle$-orbit. If $x \in A_i$ and $y \in A_j$, then, using the proof of Proposition 3.11, we see that

$$d(x,y) \geq (|i-j|-2)|\log \lambda|\,,$$

and thus, a compact subset K of $\mathbb{H}^n$ being bounded satisfies $K \subset \bigcup_{k=N}^M A_k = \mathscr{A}_K$. If $|k| \geq M-N$, then $L_\lambda^k(\mathscr{A}_K) \cap \mathscr{A}_K$ is empty, which implies that the cardinality of the set $\{g \in \langle L_\lambda \rangle : g(K) \cap K \neq \emptyset\}$ is at most $M-N$.

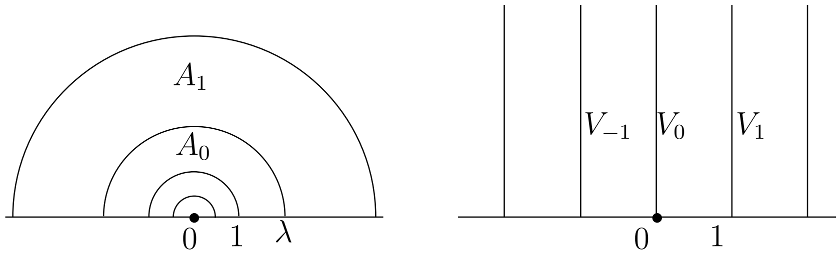

Figure 7

(2) Let $b \in \mathbb{R}^{n-1} = \mathbb{R}^n \times \{0\}$, $b \neq 0$, and let $T_b(x) = x + b$. The group $\langle T_b \rangle$ is a discontinuous group of isometries of the upper halfspace model of $\mathbb{H}^n$. The vertical strips

$$V_k = \{x \in \mathbb{H}^n : k\|b\|^2 \leq (b|x) < (k+1)\|b\|^2\}$$

are disjoint and they contain exactly one point from each $\langle T_b \rangle$-orbit. Any compact subset K of $\mathbb{H}^n$ is contained in some ball $B(a,R)$. The Euclidean radius of $B(a,r)$ is $a_n \cosh R$, which implies that K is contained in a finite union of the strips V_k, and the claim follows as in part (1).

(3) Let R_k be the rotation $R_k(z) = e^{2\pi i/k}z$ in the Poincaré disk model of the hyperbolic plane. The group generated by R_k is finite (of order k), so it is discontinuous. On the other hand, if $\alpha \in \mathbb{R} - \mathbb{Q}$, then the $\langle z \mapsto e^{2\pi i\alpha}z \rangle$-orbit of any point $z \in \mathbb{H}^2 - \{0\}$ is dense in the circle $\{w \in \mathbb{H}^2 : \|w\| = \|z\|\}$, and therefore

The action of the group G on $\mathbb{H}^n$ defines an equivalence relation $\sim$ in $\mathbb{H}^n$ by setting $x \sim y$ if and only if $Gx = Gy$. In other words, $x \sim y$ if and only if there is some $g \in G$ such that $g(x) = y$. The quotient space which is the set of G-orbits is denoted by $G\backslash\mathbb{H}^n$ (or very often $\mathbb{H}^n/G$).

Example 5.2. When $n = 2$, then the quotient spaces in Examples 5.1(1) and (2) are homeomorphic to to an annulus. Geometrically, the quotient spaces of examples (1) and (2) are very different: The proof of Proposition 3.11 implies that the image of the positive imaginary axis in $\langle L_\lambda \rangle\backslash\mathbb{H}^2$ is the shortest curve (a closed geodesic) in its free homotopy class of curves that are homotopic to the boundary components of the annulus. On the other hand, considering the images of horocycles centered at ∞ shows that there is no shortest loop in the corresponding homotopy class on $\langle T_b \rangle\backslash\mathbb{H}^2$. Note that the quotient maps $\pi\colon \mathbb{H}^2 \to \langle L_\lambda \rangle\backslash\mathbb{H}^2$ and $\pi\colon \mathbb{H}^2 \to \langle T_b \rangle\backslash\mathbb{H}^2$ are local homeomorphisms.

Theorem 5.3. *A group $G < \mathrm{Isom}(\mathbb{H}^n)$ is discontinuous if and only if all G-orbits are discrete and closed, and the stabilizers of points are finite.*

Proof. Assume first that G is discontinuous. Let $x \in \mathbb{H}^n$, and assume that y is an accumulation point of the orbit Gx. There is a sequence $(g_i)_{i\in\mathbb{N}} \in G$, such that $g_i(x) \to y$ as $i \to \infty$ with $g_i(x) \neq g_j(x)$ if $i \neq j$. The set $\{g_i(x) : i \in \mathbb{N} \cup \{0\}\} \cup \{y\}$ is compact. But now for all $k \in \mathbb{N}$, we have $g_k(x) \in g_k K \cap K$. Thus, the set $\{g \in G : gK \cap K \neq \emptyset\}$ is infinite, which is a contradiction as we assumed that G is discontinuous. Thus, the orbit Gx is discrete and closed. The fact that the stabilizer of any point is finite is clear from discontinuity because points are compact sets.

Assume now that each G-orbit is discrete and closed and that stabilizers of points are finite. Assume that the group G is not discontinuous. Then there is a compact set $K \subset \mathbb{H}^n$ such that the set $\{g \in G : gK \cap K \neq \emptyset\}$ is infinite. Thus, there is a sequence $(g_i)_{i\in\mathbb{N}} \in G$ such that $g_i \neq g_j^{\pm 1}$ when $i \neq j$ and such that $g_i K \cap K \neq \emptyset$ for all $i \in \mathbb{N}$. For each $i \in \mathbb{N}$, there is some $x_i \in K$ such that $g_i(x_i) \in K$. As K is compact, by passing to a subsequence, there are points $x, y \in K$ such that $x_i \to x$ and $y_i \to y$ as $i \to \infty$. All the maps g_i are isometries, so we have

$$d(g_j(x), y) \leq d(g_j(x), g_j(x_j)) + d(g_j(x_j), y) = d(x, x_j) + d(g_j(x_j), y) \to 0$$

as $j \to \infty$. The sequence $(g_j(x))_{j\in\mathbb{N}}$ is infinite because otherwise the stabilizer of x is infinite. Thus, y is an accumulation point of Gx, which is a contradiction. □

In fact, it is enough to check the condition of Theorem 5.3 for just one point to conclude discontinuity of the group:

Proposition 5.4. *Let $G < \mathrm{Isom}(\mathbb{H}^n)$. If the G-orbit of $x \in \mathbb{H}^n$ is discrete and closed and the stabilizer of G is finite, then the same holds for any point in $\mathbb{H}^n$. In particular, G is discontinuous.*

Proof. Assume that y_0 is an accumulation point of Gy. Then, since stabilizer of x is finite and all elements of G are isometries, the closed ball $\bar{B}(y_0, d(x, y) + 1)$ contains infinitely many points from Gx. But such a set has an accumulation point and thus, Gx cannot be discrete and closed.

Assume that the stabilizer of a point y is infinite. Then, since the stabilizer of x is finite, the $\mathrm{Stab}_G(y)$-orbit of x has an accumulation point. □

Corollary 5.5. *A group $G < \mathrm{Isom}(\mathbb{H}^n)$ is discontinuous if for some $x \in \mathbb{H}^n$, Gx is discrete and closed and the stabilizer of x is finite.* □

Let us define a function $d \colon G\backslash\mathbb{H}^n \times G\backslash\mathbb{H}^n \to \mathbb{R}$ by

$$d(Gx, Gy) = \min_{g,h\in G} d(g(x), h(y))$$

for any $Gx, Gy \in G\backslash\mathbb{H}^n$. It is easy to check that $d(Gx, Gy) = d(x, Gy) = d(Gx, y)$. The proof of the following basic result uses the facts that the G-orbits of a discontinuous group are closed and that all the maps under consideration are isometries.

Proposition 5.6. *If G is a discontinuous group of isometries of $\mathbb{H}^n$, then $(G\backslash\mathbb{H}, d)$ is a metric space.*

Proof. Let $Gx \neq Gy$. Then, $d(Gx, Gy) = d(x, Gy) > 0$ because Gy is closed and $x \notin Gy$. It is clear that d is symmetric, so it remains to check the triangle inequality. Let $Gx, Gy, Gz \in G\backslash\mathbb{H}^n$. Now,

$$d(x, g(y)) + d(y, h(z)) = d(x, g(y)) + d(g(y), gh(z)) \geq d(x, gh(z)) \,.$$

Thus,

$$d(Gx, Gz) = d(x, Gz) \leq d(x, Gy) + d(y, Gz) = d(Gx, Gy) + d(Gy, Gz) \,. \quad \square$$

If none of the elements of $G - \{\text{id}\}$ have fixed points in $\mathbb{H}^n$, then we say that G acts *without fixed points* and the action of G on $\mathbb{H}^n$ is said to be *free.*

Proposition 5.7. *If G is a discontinuous group of isometries of $\mathbb{H}^n$ that acts on $\mathbb{H}^n$ without fixed points, then the quotient map $\pi \colon \mathbb{H}^n \to G\backslash\mathbb{H}^n$, $\pi(x) = Gx$, is a locally isometric covering map.*

Proof. Let $x \in \mathbb{H}^n$, and let $s = \frac{1}{4}d(x, Gx - \{x\})$. Let us show that the restriction of π to $B(x, s)$ is an isometry. Let $y, z \in B(x, s)$. If $g \in G - \{\text{id}\}$, then

$$d(y, g(z)) \geq d(x, g(x)) - d(x, y) - d(g(x), g(z)) \geq 4s - 2s = 2s\,.$$

Thus, $d(Gx, Gy) = d(x, y)$.

The quotient map is surjective by definition, and the preimage of the open set $B(Gx, s) = \pi(B(x, s))$ is the disjoint union $\bigsqcup_{g\in G} B(g(x), s) = \bigsqcup_{g\in G} gB(x, s)$, which implies that π is a covering map. □

Example 5.8. In Examples 5.1(1) and (2), the groups act freely, and the quotient maps are locally isomorphic covering maps. For any $\lambda \neq 1$, let $X_\lambda = \langle L_\lambda\rangle\backslash\mathbb{H}^2$ with the quotient metric. The shortest homotopically nontrivial curve on X_λ has length $|\log\lambda|$. Thus, the spaces X_λ and X_μ, $\lambda, \mu > 1$ are isometric if and only if $\lambda = \mu$. On the other hand, if $Y_b = \langle T_b\rangle\backslash\mathbb{H}^2$, $b \in \mathbb{R} - \{0\}$, with the quotient metric, then Y_b is not isomorphic with any X_λ because there is no shortest loop in the free homotopy class parallel to the boundary components of the annulus.

Furthermore, all the metric spaces Y_b, $b \neq 0$ are isometric. To see this, note that for any $b \neq 0$, $\langle T_b\rangle = \langle T_{-b}\rangle$, so we only need to consider the case $b > 0$. The mapping $L_{\sqrt{b}}$ induces an isomorphism from Y_1 to Y_b: Assume that $w = T_1^k z$. Now $L_{\sqrt{b}}(w) = L_{\sqrt{b}}T_1^k L_{\sqrt{b}}^{-1} L_{\sqrt{b}}(z) = T_b^k L_{\sqrt{b}} z$, which means that the points $L_{\sqrt{b}}(z)$ and $L_{\sqrt{b}}(w)$ are equivalent under the action of the group generated by T_b. This implies that $L_{\sqrt{b}}$ induces a mapping from Y_1 to Y_b by the rule $L_{\sqrt{b}}(\langle T_1\rangle x) = \langle T_b\rangle x$. Clearly, $L_{\sqrt{b}}^{-1}$ induces its inverse mapping. The induced mapping is an isometry by Proposition 5.7.

In example 5.1(3), the rotations R_k have a fixed point at 0, and the quotient map is not a covering map and it is not locally isometric in any neighbourhood of 0. In this case, the quotient space is a *cone.*

If $G < \text{Isom}(\mathbb{H}^n)$ acts on $\mathbb{H}^n$ freely and discontinuously, then the quotient space $G\backslash\mathbb{H}^n$ is a *hyperbolic manifold*: Each point Gx has a small neighbourhood given by the proof of Proposition 5.7 which is isometric with a ball in $\mathbb{H}^n$.

A subset $F \subset \mathbb{H}^n$ is a *fundamental set* of $G < \text{Isom}(\mathbb{H}^n)$ if it contains exactly one point from each G-orbit. A nonempty open subset $D \subset \mathbb{H}^n$ is a *fundamental domain* of G if

- the closure of D contains a fundamental set of G,
- $gD \cap D = \emptyset$ for all $g \in G - \{\text{id}\}$, and
- the boundary of D has measure 0.

Proposition 5.9. *If $G < \mathrm{Isom}(\mathbb{H}^n)$ has a fundamental domain, then G is discontinuous.*

Proof. The orbit of any point in the fundamental domain is discrete and closed, and the second condition in the definition of a fundamental domain implies that the stabilizer of such a point consists of the identity. The claim follows from Proposition 5.4. □

Let $G < \mathrm{Isom}(\mathbb{H}^n)$ be a discontinuous group, and let $x \in \mathbb{H}^n$ be a point that is not fixed by any element of g. The set

$$D(x) = \{y \in \mathbb{H}^n : d(x, y) < d(g(x), y)\}$$

is a *Dirichlet domain* of G *centered* at x.

Proposition 5.10. *Let $G < \mathrm{Isom}(\mathbb{H}^n)$ be a discontinuous group and let $w \in \mathbb{H}^n$. Then G has a Dirichlet domain centered at w. The Dirichlet domain $D(w)$ is a locally finite polytope which is a fundamental domain of G.*

Proof. Note that if we set for any $g \in G$,

$$H_g(w) = \{x \in \mathbb{H}^n : d(x, w) < d(x, g(w))\},$$

then each H_g is a hyperbolic halfspace and

$$D(w) = \bigcap_{g \in G - \{\mathrm{id}\}} H_g(w).$$

The collection $(H_g)_{g \in G}$ is locally finite because $d(w, \partial H_g(w)) = \frac{1}{2} d(w, g(w))$ and therefore, by discontinuity, any ball centered at w intersects only a finite number of the collection of hyperplanes $(\partial H_g)_{g \in G}$. Local finiteness implies that $D(w)$ is open and the boundary of $D(w)$ is contained in the measure zero subset $\bigcup_{g \in G} \partial H_g$ of $\mathbb{H}^n$.

Let us show that G has a fundamental set F that satisfies $D(w) \subset F \subset \overline{D(w)}$. Discreteness and closedness of the orbit Gz imply that there is some $z^* \in Gz$ for which $d(z^*, w) = d(Gz, w)$. Let us choose one such z^* for each orbit, and let

$$F = \{z^* : Gz \in G \backslash \mathbb{H}^n\}.$$

Clearly, F is a fundamental set that contains $D(w)$. Furthermore, if the open interval $]w, z^*[$ intersects some H_g, $g \in G - \{\mathrm{id}\}$, then, by definition of the halfspace H_g, we have

$$d(g^{-1}(z^*), w) = d(z^*, g(w)) < d(z^*, w),$$

which is a contradiction. Thus, $F \subset \overline{D(w)}$. □

Note that the halfspace H_g in the above proof is one the component of the complement of the bisector $\mathrm{bis}(w, g(w))$ that contains w.

We conclude with an important example of a discontinuous group of isometries of $\mathbb{H}^2$.

Example 5.11. The classical modular group $\Gamma = \mathrm{SL}_2(\mathbb{Z})$ acts by isometries on $\mathbb{H}^2$. We will show that the action is discontinuous and we will study the quotient space. Note that if $\mathrm{Im}\, z > 1$, then z is not a fixed point of any elliptic element of Γ: If $z = Az = \frac{az+b}{cz+d}$, then, since the coefficients a, b, c, d are real,

$$\mathrm{Im}\, z = \frac{\mathrm{Im}\, z}{|cz + d|^2}.$$

If $c \neq 0$, then $|cz + d|^2 > |c||\mathrm{Im}\, z| \geq |\mathrm{Im}\, z| > 1$, but this is not possible. Thus, the only possibility is $c = 0$ and $d = \pm 1$, which implies $z = Az = z + b$. But such a mapping has a fixed point in $\mathbb{H}^2$ if and only if $b = 0$. Thus, the only elements of Γ that fix z are $\pm I_2$.

Let $Az = z + 1$ and $Bz = -\frac{1}{z}$. We will show that for any $t > 1$, the Dirichlet domain centered at ti has a simple form:

$$D(ti) = P = H_A \cap H_{A^{-1}} \cap H_B = \{z \in \mathbb{H}^2 : -\frac{1}{2} < \operatorname{Re} z < \frac{1}{2},\ |z| > 1\}\,.$$

The inclusion $D(ti) \subset P$ is immediate from the definition of $D(ti)$.

To prove the other inclusion, let $z \in P$ and let $C = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \neq I_2$ such that $Cz \in P$. If $c = 0$, then $a = d = \pm 1$ and $b \neq 0$. This implies

$$|\operatorname{Re} Cz| = |\operatorname{Re} z + b| \geq |b| - |\operatorname{Re} z| > |b| - \frac{1}{2} \geq \frac{1}{2}\,,$$

so that $Cz \notin P$. Thus, we must have $c \neq 0$. Now, since $|z| > 1$ and $|\operatorname{Re} z| > \frac{1}{2}$, we have

$$|cz + d|^2 = c^2|z|^2 + 2cd \operatorname{Re} z + d^2 > c^2 + d^2 - |cd| = (|c| - |d|)^2 - |cd| > 1\,.$$

This implies $\operatorname{Im} Cz < \operatorname{Im} z$. Applying this observation to the point Cz and the transformation C^{-1} gives the opposite inequality $\operatorname{Im} Cz < \operatorname{Im} z$, which is a contradiction.

It is best to think of P as a 4-sided polygon with vertices at ∞, $\rho = e^{\pi i/3}$, i and $-\bar{\rho}$ with angles 0, $\frac{\pi}{3}$, π, and $\frac{\pi}{3}$, respectively, at the vertices. The transformation A identifies the vertical sides and B identifies the two sides on the Euclidean unit circle. The quotient space $\Gamma\backslash\mathbb{H}^2$ is homeomorphic to a 2-plane. Geometrically, it is locally isometric with the hyperbolic plane at all points except at Γi and $\Gamma\rho = \Gamma(-\bar{\rho})$, where it has cone points of angles π and $\frac{2\pi}{3}$, respectively. Using Proposition 3.18 , we see that the area of $\Gamma\backslash\mathbb{H}^2$ is $\pi/3$.

The classical modular group has a number of important subgroups that act on $\mathbb{H}^2$ without fixed points, for example, the *level p principal congruence subgroup*

$$\Gamma(p) = \{\gamma \in \Gamma : \gamma \equiv \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \mod p\}$$

and the *commutator subgroup* $[\Gamma, \Gamma]$ generated by all elements of the form $\gamma_1\gamma_2\gamma_1^{-1}\gamma_2^{-1}$ with $\gamma_1, \gamma_2 \in \Gamma$.

References

[Ahl] L. V. Ahlfors. *Complex analysis.* McGraw-Hill Book Co., New York, third edition, 1978.

[And] J. W. Anderson. *Hyperbolic geometry.* Springer Undergraduate Mathematics Series. Springer-Verlag London Ltd., London, second edition, 2005.

[Bea] A. F. Beardon. *The geometry of discrete groups*, volume 91 of *Graduate Texts in Mathematics.* Springer-Verlag, New York, 1983.

[BH] M. R. Bridson and A. Haefliger. *Metric spaces of non-positive curvature*, volume 319 of *Grundlehren der Mathematischen Wissenschaften.* Springer-Verlag, Berlin, 1999.

[Gre] W. Greub. *Linear algebra.* Springer-Verlag, New York, fourth edition, 1975. Graduate Texts in Mathematics, No. 23.

Suggestions for reading material

- Sections I.2 and I.6 of [BH] contain material on the various models of hyperbolic space. A different approach to the hyperbolic plane can be found in [And] and Chapter 7 of [Bea].
- Hyperbolic trigonometry is treated extensively in the second half of Chapter 7 of [Bea].