NOTES ON HYPERBOLIC GEOMETRY

KYLE ORMSBY

Caveat emptor — These notes are in draft form and will evolve throughout the semester. Please contact the author at ormsbyk@reed.edu with any comments or corrections.

Much of these notes are highly parallel to Birger Iversen's *Hyperbolic geometry* [Ive92] and they should not be considered original work.

CONTENTS

1. A SURPLUS OF TRIANGLES

This activity is inspired by Kathryn Mann's notes *DIY hyperbolic geometry* [Man15]. Jay Ewing kindly provided the class with approximately 900 equilateral triangles cut from 100# card stock by the College's laser cutter.

Take a stack of triangles and start taping them together along edges, with the restriction that exactly n triangles meet at each vertex, where n is some fixed positive integer. You should find some familiar shapes for $n = 3, 4, 5$: the tetrahedron, octahedron, and icosahedron.

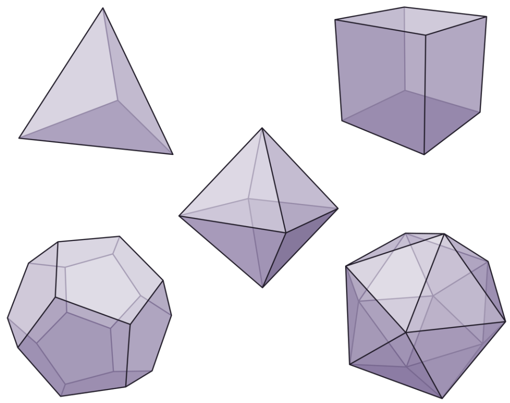

FIGURE 1. The Platonic solids. You should have built the tetrahedron, octahedron, and icosahedron. The cube and dodecahedron are built from squares and pentagons, respectively. Image: Wikimedia Commons.

With $n = 6$, your job remains easy, but you now produce an infinite, flat triangular lattice.

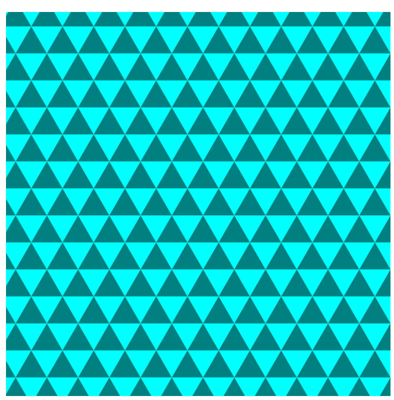

FIGURE 2. The infinite triangular lattice produced by placing six triangles at each vertex. Image: Wikimedia Commons.

When $n = 7$, life becomes more interesting. With $420°$ worth of angle at each vertex, the arrangement cannot possibly fit in the plane. The shape you are creating is an approximation of the *hyperbolic plane*.

A *geodesic* in a metric space is a distance-minimizing curve. On the polygonal hyperbolic plane, these are "straight lines," where 'straight' has the obvious meaning on flat triangles, and lines continue straight over edges by temporarily flattening them. (Geodesics through vertices are problematic.)

Excercise 1.1. Draw several geodesics on your polygonal hyperbolic plane and try to describe their features. More specifically:

(a) Draw two geodesics that start off parallel (inside of a triangle) but eventually diverge.
(b) Draw a geodesic that starts close to and parallel to an edge of a triangle. Observe that if follows a path made up of triangle edges. Call such a path an *edge path geodesic*. Try to describe all edge-path geodesics.
(c) Draw a large (not inside a paper triangle) triangle made up of geodesics.
(d) Is it possible to draw a large rectangle (quadrilateral where all sides meet at right angles)?

We will now investigate some properties of area in our polygonal hyperbolic plane.

Definition 1.2. A *polygonal disk of radius* r consists of all triangles with vertices at most r edges from a central vertex. The *combinatorial area* of a region made up of triangles is the number of triangles in the region. The *combinatorial circumference* of a region made up of triangles is the number of triangle edges on the boundary of the region.

Excercise 1.3. Find formulæ for the combinatorial area and circumference of a polygonal disk of radius r in the Euclidean plane (six triangles per vertex).

Excercise 1.4. Find formulæ or estimates for the combinatorial area and circumference of a polygonal disk of radius r in the polygonal hyperbolic plane.

You may find Figure 3 useful in answering Excercise 1.4.

2. What is hyperbolic geometry?

Here are some slightly non-traditional axioms for Euclidean geometry. Throughout, we assume that X is a nonempty metric space.[1] A *line* (or *geodesic line*) in X is a distance-preserving function $\gamma\colon \mathbb{R} \to X$, where $\mathbb{R}$ has the standard metric. Two lines are *parallel* if their images are disjoint. An *isometry* is a function between metric spaces $\sigma\colon X \to Y$ which preserves distances and has a distance-preserving inverse.

Incidence axiom: A unique line passes through any two distinct points of X.

Reflection axiom: If $\gamma\colon \mathbb{R} \to X$ is a line in X, then $X \smallsetminus \gamma(\mathbb{R})$ has two connected components. There exists an isometry $\sigma\colon X \to X$ which fixes $\gamma(\mathbb{R})$ pointwise but interchanges the connected components of $X \smallsetminus \gamma(\mathbb{R})$.

Parallel axiom: Given a line γ in X and a point $p \in X \smallsetminus \gamma(\mathbb{R})$, there is a unique line through p parallel to γ.

[1] Recall that a *metric space* is a set X equipped with a *metric*, that is, a function $d\colon X \times X \to \mathbb{R}$ such that for all $x, y, z \in X$, $d(x,y) = 0 \iff x = y$ (identity of indiscernibles), $d(x,y) = d(y,x)$ (symmetry), and $d(x,z) \leq d(x,y) + d(y,z)$ (triangle inequality).

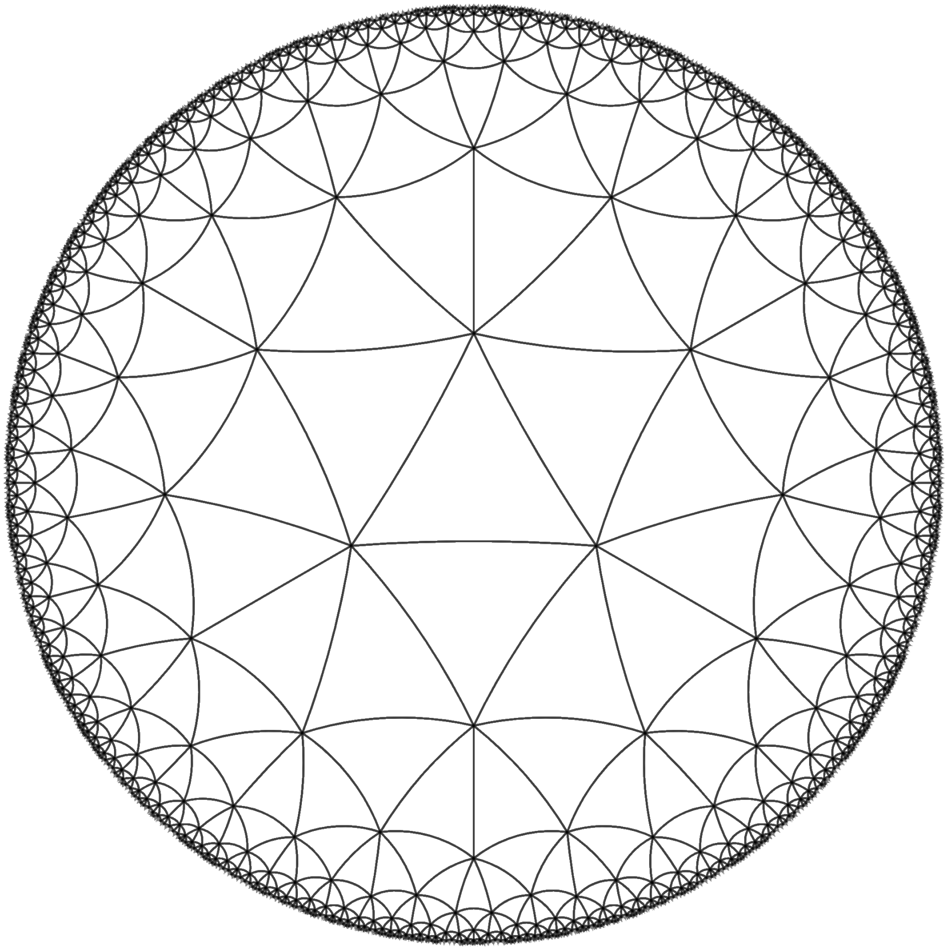

FIGURE 3. A tessellation of the Poincaré disk model of the hyperbolic plane with seven equilateral triangles around each vertex. Created with the utility Make hyperbolic tilings of images.

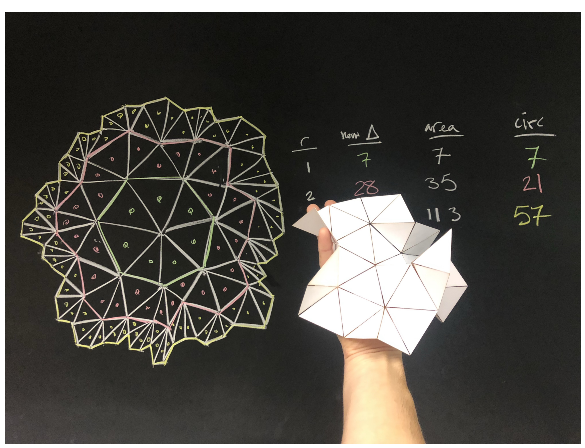

FIGURE 4. The author holding a paper model of a polygonal disk of radius 2 in front of some preliminary data on Excercise 1.4.

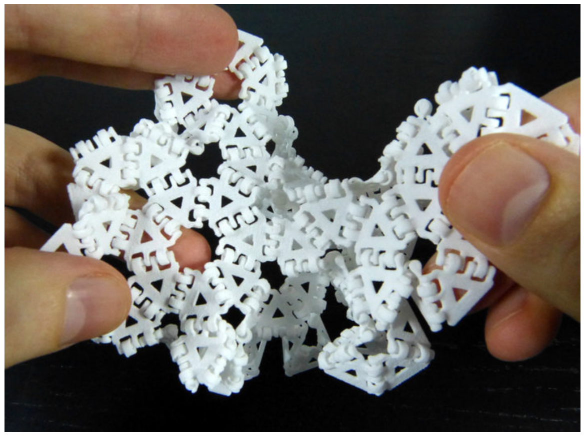

FIGURE 5. Henry Segerman has made some delightful 3D printed objects representing the polygonal hyperbolic plane. See this Numberphile video for a live demonstration.

Theorem 2.1. *A metric space which satisfies all three of these axioms is isometric to the Euclidean plane. A space which satisfies the first two axioms but not the third is isometric to the hyperbolic plane $\mathbb{H}^2$ (after rescaling).*

We will not study the proof of this theorem, but rather state it to motivate the relative universality of the hyperbolic plane.

Remark 2.2. Spherical geometry is another interesting system, but it fails to satisfy the incidence axiom: two antipodal points have infinitely many lines (great circles) passing through them.

The Poincaré half plane. The French polymath Henri Poincaré (1854–1912) is one of the key contributors to the theoretical underpinnings of hyperbolic geometry. He introduced two models of $\mathbb{H}^2$, one based on the upper half plane in $\mathbb{C}$, and the other based on the unit disk. He also supplied one of the most memorable quotations regarding the process of mathematical discovery:

> For fifteen days I strove to prove that there could not be any functions like those I have since called Fuchsian functions.[2] I was then very ignorant; every day I seated myself at my work table, stayed an hour or two, tried a great number of combinations and reached no results. One evening, contrary to my custom, I drank black coffee and could not sleep. Ideas rose in crowds; I felt them collide until pairs interlocked, so to speak, making a stable combination. By the next morning I had established the existence of a class of Fuchsian functions, those which come from the hypergeometric series; I had only to write out the results, which took but a few hours.

The Poincaré upper half plane H is the set

$$H = \{z \in \mathbb{C} \mid \operatorname{Im} z > 0\}$$

equipped with the metric

$$d(z, w) = \operatorname{arccosh}\left(1 + \frac{|z - w|^2}{2 \operatorname{Im}(z) \operatorname{Im}(w)}\right).$$

[2]These are analytic functions invariant under the action of a Fuchsian group. Fuchsian groups will be one of our main topics of study and will be defined in due course.

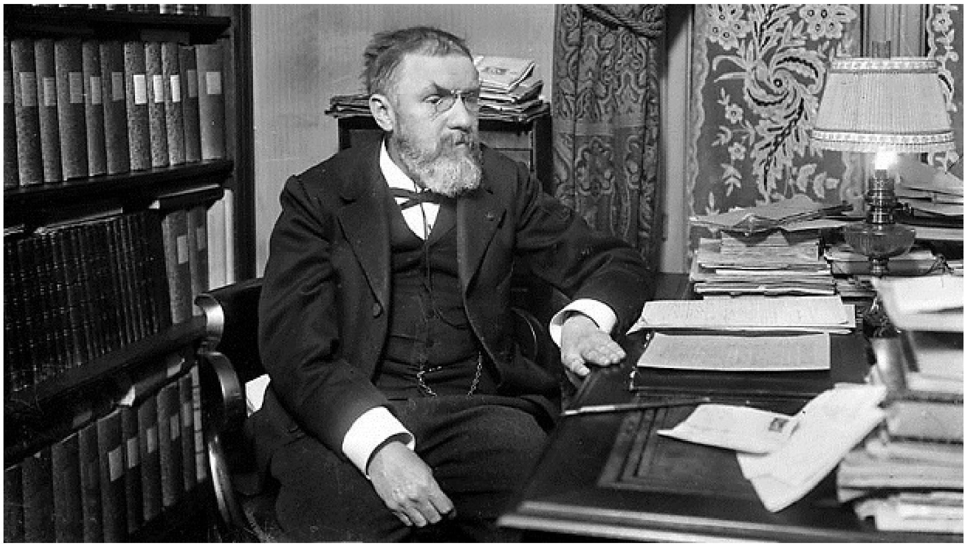

FIGURE 6. Henri Poincaré, a central figure in late nineteenth century geometry, topology, and physics. Image: https://famous-mathematicians.com/henri-poincare/

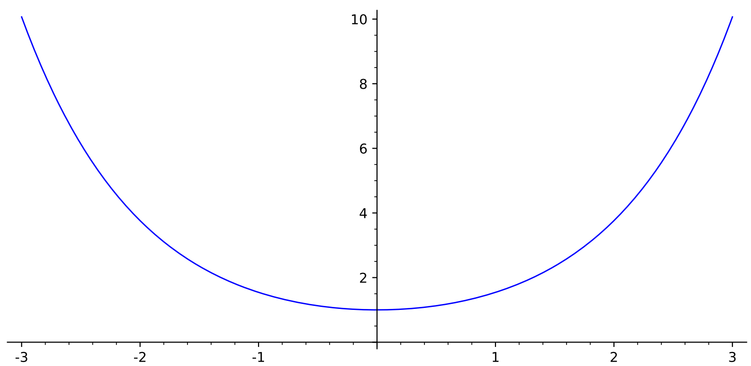

FIGURE 7. The plot of cosh.

Recall that the *hyperbolic cosine* function is

$$\cosh(x) = \frac{e^x + e^{-x}}{2}$$

with plot shown in Figure 7.

The function arccosh is the inverse of cosh restricted to the nonnegative reals.

The lines in H are Euclidean circles with centers on the real axis or Euclidean lines perpendicular to the real axis. The reader is invited to think about how this model satisfies the incidence and reflection axioms, but not the parallel axiom, perhaps by consulting Figure 8.

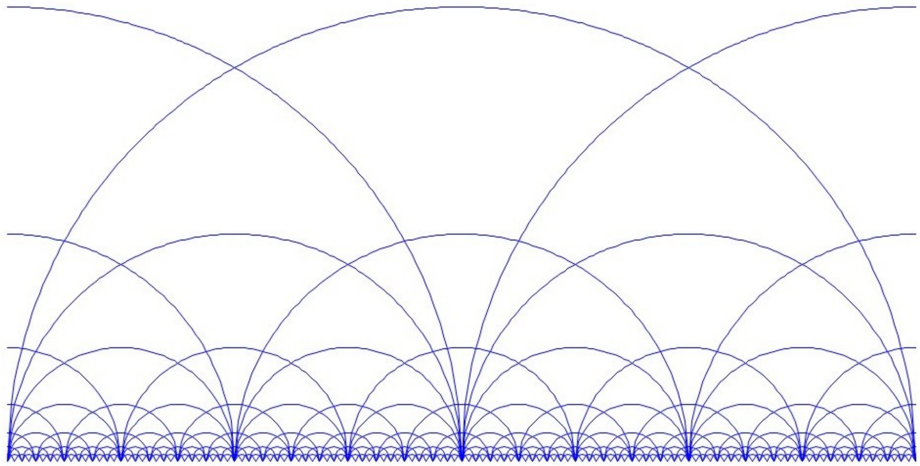

FIGURE 8. Some not-so-random geodesics in H. Image: https://thatsmaths.com/2013/10/11/poincares-half-plane-model/.

The group $\mathrm{SL}_2(\mathbb{R})$ of 2×2 real matrices with determinant one acts on H via isometries via the assignment

$$z \longmapsto \frac{az+b}{cz+d} \qquad \text{for} \qquad \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{R}).$$

In particular, $\mathrm{SL}_2(\mathbb{Z})$ acts on H via isometries, and this is crucial to interactions between hyperbolic geometry and number theory.

The Poincaré disk. The open unit disk[3] $D = \{z \in \mathbb{C} \mid |z| < 1\}$ can also be used as a model for $\mathbb{H}^2$. The metric again has an expansion in terms of an inverse hyperbolic trig function, and the geodesics are equally easy to describe: they are Euclidean circles and lines orthogonal to the unit circle $S^1 = \partial D$. See Figure 3 for an aesthetically pleasing arrangements of geodesics in D.

Discrete groups. Reflections in the lines of the tessellation in Figure 3 generate a discrete group of isometries of $\mathbb{H}^2$. Starting with a discrete group Γ of isometries of $\mathbb{H}^2$, one may construct a polygon Δ with a side pairing (called a *fundamental domain*) so that $\mathbb{H}^2/\Gamma$ is reconstructed by gluing together the prescribed sides of Δ.

Poincaré was concerned with the converse problem of determining when a given hyperbolic polygon with side pairing generates a discrete group with the given polygon as fundamental domain. The necessary and sufficient conditions he deduced are called *Poincaré's Theorem*, and the elucidation of its statement and proof will be one of our primary concerns this semester.

When Γ is a discrete of $\mathrm{Isom}(\mathbb{H}^2)$, we call it a *Fuchsian group*. The quotient spaces $\mathbb{H}^2/\Gamma$ does not satisfy the incidence and reflection axioms, but it does *locally* satisfy them; in other words, $\mathbb{H}^2/\Gamma$ is locally isometric to $\mathbb{H}^2$ and as such is a geometry called a *hyperbolic surface*. It turns out that every hyperbolic surface is isometric to some $\mathbb{H}^2/\Gamma$, Γ a Fuchsian group, highlighting the importance of these objects.

[3]Or is it disc? The spelling 'disk' seems to be preferred in American English, unless one is interacting with antiquated compact disc (CD) technology.

FIGURE 9. Poincaré named the discrete subgroups of $\mathrm{Isom}(\mathbb{H}^2)$ after Lazarus Fuchs (1833–1902).

Sins of ommission. Following Iversen [Ive92], we will take a metric and linear algebraic approach to hyperbolic geometry, largely forgoing differential and Riemannian geometry (including curvature).

3. QUADRATIC FORMS

Students from last semester's Math 412 must bite their tongues during the following two lectures.

Orthogonality. Let k be a field of characteristic different from 2, and let E be a k-vector space of finite dimension n. A *quadratic form* on E is a function $Q\colon E \to \mathsf{k}$ satisfying $Q(\lambda v) = \lambda^2 Q(v)$ for all $\lambda \in \mathsf{k}$, $v \in E$, and such that its *polarization* $\langle\ ,\ \rangle = \langle\ ,\ \rangle_Q : E \times E \to \mathsf{k}$ given by

$$\begin{aligned} \langle\ ,\ \rangle : E \times E &\longrightarrow \mathsf{k} \\ (v, w) &\longmapsto \frac{1}{2}(Q(v+w) - Q(v) - Q(w)) \end{aligned}$$

is a symmetric bilinear form. Note that $\operatorname{char} \mathsf{k} \neq 2$ is essential for defining the polarization.

Excercise 3.1. If $k = \mathbb{R}$, $E = \mathbb{R}^n$, and $Q\colon \mathbb{R}^n \to \mathbb{R}$ is twice-differentiable and satisfies $Q(\lambda v) = \lambda^2 v$, then the polarization of Q is necessarily a symmetric bilinear form.

Proposition 3.2. Polarization is a bijective correspondence between quadratic forms on E and symmetic bilinear forms on E.

Proof. We claim that the assignment $\langle\ ,\ \rangle \longmapsto (v \mapsto \langle v, v\rangle)$ is a two-sided inverse to polarization. Note that $\langle \lambda v, \lambda v\rangle = \lambda^2 \langle v, v\rangle$, so this map is in fact quadratic. The computation

$$\frac{1}{2}(Q(2v) - Q(v) - Q(v)) = \frac{1}{2}(4Q(v) - 2Q(v)) = Q(v)$$

show that a quadratic form may be recovered from its polarization. We leave the other direction as an exercise for the reader. □

If a quadratic/symmetric bilinear form is fixed, we call $Q(v) = \langle v, v\rangle$ the *norm* of v (with respect to the fixed quadratic/symmetric bilinear form). Two vectors $v, w \in E$ are *orthogonal* when $\langle v, w\rangle = 0$. Two linear subspaces $V, W \leq E$ are orthogonal when $\langle v, w\rangle = 0$ for all $v \in V$, $w \in W$. Given a linear subspace $V \leq E$, we write

$$V^{\perp} = \{w \in E \mid \langle v, w\rangle = 0 \text{ for all } v \in V\}$$

and call this space the *orthogonal complement* of V.

Excercise 3.3. Check that $V^{\perp}$ is a linear subspace of E.

Definition 3.4. A quadratic form Q with polarization $\langle\ ,\ \rangle$ on a finite-dimensional vector space E is called *nonsingular* if $\langle v, w\rangle = 0$ for all $v \in E$ implies that $w = 0$.

Excercise 3.5. Given a symmetric bilinear form $\langle\ ,\ \rangle$ on E, there is an associated linear transformation

$$\begin{aligned} q\colon E &\longrightarrow E^* \\ v &\longmapsto (w \mapsto \langle v, w\rangle) \end{aligned}$$

where E^* is the k-linear dual [4] of E. Show that $\langle\ ,\ \rangle$ is nonsingular if and only if the associated map $E \to E^*$ is an isomorphism.

Theorem 3.7. *Let E be a finite-dimensional k-vector space equipped with a* nonsingular *quadratic form Q. Then for any linear subspace $V \leq E$, we have*

$$\dim V + \dim V^{\perp} = \dim E.$$

Proof. By Excercise 3.5, the map $q\colon E \to E^*$ is an isomorphism. The inclusion map $i\colon V \hookrightarrow E$ has dual

$$\begin{aligned} i^*\colon E^* &\longrightarrow V^* \\ f &\longmapsto f \circ i. \end{aligned}$$

We claim that i^* is surjective. To wit, if $v_1, \ldots, v_k$ is a basis of V, we may extend it to a basis $v_1, \ldots, v_n$ of E. The reader may check that the matrix for i^* with respect to bases $v_1^*, \ldots, v_n^*$ and $v_1^*, \ldots, v_k^*$ is $(I_k \mid 0)$, so i^* is surjective.

---

[4] Dual vector spaces play an important role in linear algebra and an outsized one in the theory of symmetric bilinear and quadratic forms.

Definition 3.6. The k-linear *dual* of a k-vector space V is the Hom space

$$V^* = \mathrm{Hom}_{\mathsf{k}}(V, \mathsf{k}).$$

Elements of V^* are called *linear functionals* or *dual vectors*.

There is a canonical map

$$\begin{aligned} V &\longrightarrow (V^*)^* \\ v &\longmapsto (f \mapsto f(v)) \end{aligned}$$

which is always injective and is an isomorphism when V is finite-dimensional. (We call the map *canonical* because it does not depend on the choice of a basis or coordinates.) For V finite-dimensional, it is also the case that $V \cong V^*$, but this isomorphism in non-canonical. Indeed, after choosing an ordered basis $\{v_1, \ldots, v_n\}$ of V, we create a *dual basis* $\{v_1^*, \ldots, v_n^*\}$ where $v_i^*(v_j)$ is either 1 or 0 depending on whether $j = i$ or $j \neq i$. It is straightforward to prove that $\{v_1^*, \ldots, v_n^*\}$ is a basis of V^* and the linear map taking v_i to v_i^* is an isomorphism.

Given a linear transformation $f\colon V \to W$, we can form the *dual transformation* $f^* : W^* \to V^*$ which takes $g : W \to \mathsf{k}$ to the composite linear functional $g \circ f$. This defines an injective linear transformation

$$\mathrm{Hom}_{\mathsf{k}}(V, W) \longrightarrow \mathrm{Hom}_{\mathsf{k}}(W^*, V^*)$$

which is an isomorphism when both vector spaces are finite-dimensional. (Check this!)

Now consider the the composite linear map $i^* \circ q \colon E \to V^*$ which is also surjective (since it is the composite of an isomorphism and then a surjection). By rank-nullity,

$$\dim E = \dim V^* + \dim \ker(i^* \circ q).$$

Finally, one may check that $\ker(i^* \circ q) = V^\perp$, and we already know $\dim V^* = \dim V$, so we deduce the desired equality. □

Given a quadratic form Q on a vector space E with $\langle\ ,\ \rangle = \langle\ ,\ \rangle_Q$, one may form a *Gram matrix* for Q by choosing a basis $v_1, \ldots, v_n$ of E and then creating $G \in \mathrm{Mat}_{n\times n}(\mathsf{k})$ with

$$G_{ij} = \langle v_i, v_j \rangle .$$

Note that G is a symmetric matrix. Also note that G is the matrix for $q \colon E \to E^*$ with respect to the bases $v_1, \ldots, v_n$ and $v_1^*, \ldots, v_n^*$.

Excercise 3.8. Show that Q is nonsingular if and only if the Gram matrix G of Q has $\det G \neq 0$.

Let (E, Q) be a vector space with quadratic form Q. An *isometry* of (E, Q) with another such pair (E', Q') is a linear isomorphism $\sigma \colon E \to E'$ such that $Q = Q' \circ \sigma$. The *automorphisms* of (E, Q) are the isometries $(E, Q) \to (E, Q)$. Under composition, automorphisms form a group $\mathrm{O}(Q)$ — or $\mathrm{O}(E)$ when no confusion is possible — called the *orthogonal group* of (E, Q).

Proposition 3.9. Let Q be a nonsingular form on E. If $\sigma \in \mathrm{O}(Q)$, then $\det \sigma = \pm 1$.

Proof. Pick a basis $v_1, \ldots, v_n$ for E and let A be the matrix for σ with respect to this basis. Then

$$\begin{aligned} \langle \sigma(v_i), \sigma(v_j) \rangle &= \left\langle \sum_k A_{ki} v_k, \sum_h A_{hj} v_h \right\rangle \\ &= \sum_{h,k} A^\top_{ik} \langle v_k, v_h \rangle A_{hj} \end{aligned}$$

which show that $A^\top G A$ is the Gram matrix for $Q \circ \sigma$ (with respect to $v_1, \ldots, v_n$). In particular, $\sigma \in \mathrm{O}(Q)$ if and only if $A^\top G A = G$. Applying the determinant, we get

$$\det G = \det A^\top \det G \det A = (\det A)^2 \det G,$$

whence $\det A = \pm 1$. □

When $Q(x) = \sum x_i^2$ for $x = (x_1, \ldots, x_n) \in \mathsf{k}^n$, we write $\mathrm{O}_n(\mathsf{k})$ for $\mathrm{O}(Q)$. Note that Q has Gram matrix I_n, so $A \in \mathrm{O}_n(\mathsf{k})$ if and only if $A^\top A = I_n$.

Theorem 3.10. *Let Q be a nonsingular quadratic form on a vector space E over $\mathbb{C}$ (or any algebraically closed field). Then there exists a basis of E with respect to which Q has Gram matrix I_n.*

Proof. See [Ive92, p.5]. □

Witt's theorem. Fix a field k with $\mathrm{char}\,\mathsf{k} \neq 2$, let E be a finite-dimensional k-vector space, and let Q be a quadratic form on E. Set $\langle\ ,\ \rangle = \langle\ ,\ \rangle_Q$. A vector $v \in E$ is *isotropic* if $Q(v) = 0$; otherwise, v is *nonisotropic*. Given a nonisotropic vector v, define a linear transformation

$$\begin{aligned} \tau_v \colon E &\longrightarrow E \\ x &\longmapsto x - 2\frac{\langle x, v \rangle}{\langle v, v \rangle} v. \end{aligned}$$

The reader may check that

$$\langle \tau_v(x), \tau_v(x) \rangle = \langle x, x \rangle ,$$

so $\tau_v \in \mathrm{O}(Q)$. The orthogonal transformation τ_v is called *reflection along* v. Note that $\tau_v|_{v^\perp} = \mathrm{id}$, while $\tau_v(v) = -v$. Hence τ_v is an *involution*: $\tau_v \circ \tau_v = \mathrm{id}$, but $\tau_v \neq \mathrm{id}$. Also note that $\det \tau_v = -1$.

We say that a group G acts on a set X when there is a function $G \times X \to X$, $(g, x) \mapsto g \cdot x$ such that $e \cdot x = x$ and $g \cdot (h \cdot x) = (gh) \cdot x$. A group action is *transitive* when for all $x, y \in X$, there exists $g \in G$ such that $g \cdot x = y$. The following proposition says that $\mathrm{O}(Q)$ acts transitively on any "Q-sphere" in E.

Proposition 3.11. For any $\lambda \in \mathsf{k}^\times$, $\mathrm{O}(Q)$ acts transitively on the set

$$S_Q(\lambda) = \{v \in E \mid Q(v) = \lambda\}.$$

Proof. Given $\sigma \in \mathrm{O}(Q)$ and $v \in S_Q(\lambda)$, we define $\sigma \cdot v = \sigma(v)$. Since σ is an isometry, $Q(\sigma(v)) = Q(v) = \lambda$, so $\sigma \cdot v \in S_Q(\lambda)$. The other properties of a group action are obvious.

To prove transitivity, suppose that $v, w \in S_Q(\lambda)$. These vectors are nonisotropic since $Q(v) = Q(w) = \lambda \neq 0$. Also observe that

$$\langle v - w, v + w\rangle = \langle v, v\rangle - \langle w, w\rangle = \lambda - \lambda = 0,$$

that is, $v - w$ and $v + w$ are orthogonal. Note then that $v = \frac{1}{2}((v - w) + (v + w))$, and thus $\lambda = Q(v) = \frac{1}{4}\langle (v - w) + (v + w), (v - w) + (v + w)\rangle = Q(v - w) + Q(v + w)$. Since $\lambda \neq 0$, we conclude that we cannot have both $Q(v - w)$ and $Q(v + w)$ equal to 0.

If $v - w$ is nonisotropic, then

$$\tau_{v-w}(v - w) = w - v \qquad \text{and} \qquad \tau_{v-w}(v + w) = v + w.$$

Adding these formulæ gives $\tau_{v-w}(v) = w$. A similar computation shows that if $v + w$ is nonisotropic then $\tau_{v+w}(v) = -w$, whence $\tau_w \circ \tau_{v+w}$ takes v to w. We conclude that the subgroup of $\mathrm{O}(Q)$ generated by reflections acts transitively on $S_Q(\lambda)$. ☐

Remark 3.12. Cartan's theorem — which we will not prove — implies that $\mathrm{O}(Q)$ is generated by reflections.

Theorem 3.13 (Witt)**.** *Let (E, Q) be a nonsingular quadratic form. Any isometry $\sigma\colon (U, Q) \cong (V, Q)$, where U, V are linear subspaces of E, can be extended to an orthogonal transformation of (E, Q).*

Proof. First suppose that U is nonsingular and proceed by induction on $\dim U$. If $\dim U = 1$, pick $0 \neq u \in U$ and set $\lambda = Q(u)$. Since $Q(\sigma(u)) = \lambda$, we can apply Proposition 3.11 find $\sigma' \in \mathrm{O}(Q)$ such that $\sigma'(u) = \sigma(u)$. It is easy to check that $\sigma'|_U = \sigma$, so σ' is our desired extension.

For the induction step, pick a nonisotropic vector $u \in U$. Use Proposition 3.11 to choose $\tau \in \mathrm{O}(Q)$ such that $\sigma(e) = \tau(e)$. It suffices to extend $\tau^{-1} \circ \sigma\colon U \to E$ to an orthogonal transformation of E. This means that, without loss of generality, it suffices to extend $\sigma\colon U \to E$ under the assumption that σ fixes a nonisotropic vector $u \in U$. Let W be the ortgonal space to u in U, and let V be the orthogonal space to u in E. Then W and V are nonsingular, $W \leq V$, and $\sigma(W) \leq V$. By the induction hypothesis, $\sigma|_W\colon W \to V$ extends to an orthogonal transformation of V, and this extension may be extended to E via the identity, completing our proof in the case where U is nonsingular.

We leave the case of U singular as a reading exercise; see [Ive92, pp.7–8]. ☐

4. Real quadratic forms

We now specialize to $\mathsf{k} = \mathbb{R}$, the field of real numbers.

FIGURE 10. Ernst Witt (1911–91) was a leading algebraist in the 20th century and founded the theory of quadratic forms over an arbitrary field. Image: Oberwolfach Photo Collection.

Sylvester types. The isometry type of a real quadratic form is determined by two numbers, called the *Sylvester type*. In order to develop this theory, we first generalize the notion of orthonormal basis.

Proposition 4.1. Let (E, Q) be a quadratic form on a real vector space E of dimension n. There exists a basis $e_1, \ldots, e_n$ for E such that

$$\langle e_i, e_j \rangle = \begin{cases} 0 & \text{if } i \neq j, \\ \pm 1 \text{ or } 0 & \text{if } i = j. \end{cases}$$

Proof. We proceed by induction on $n = \dim E$. If $Q = 0$, any basis for E satisfies these properties. Thus we may assume there exists $v \in E$ which is nonisotropic. Write $Q(v) = \varepsilon\lambda^2$ with $\varepsilon = \pm 1$ and $\lambda \in \mathbb{R}^\times$. Set $e_1 = \lambda^{-1}v$ to get $Q(e_1) = \pm 1$. This covers the $n = 1$ case. For $n > 1$, invoke the induction hypothesis to find $e_2, \ldots, e_n$ a basis of the desired type for $e_1^\perp$. The basis $e_1, \ldots, e_n$ meets our requirements. □

Definition 4.2. Call a basis of the type guaranteed by Proposition 4.1 an *orthonormal basis*.

The careful reader will note that this is a larger class of bases than those usually termed orthonormal (for which $\langle e_i, e_j \rangle = \delta_{ij}$).

Theorem 4.3 (Sylvester). *Let $e_1, \ldots, e_n$ be an orthonormal basis for (E, Q). The numbers*

$$p = \#\{i \mid \langle e_i, e_i \rangle = -1\} \qquad \textit{and} \qquad q = \#\{i \mid \langle e_i, e_i \rangle = 1\}$$

are independent of the orthonormal basis considered.

Proof. Fix an orthonormal basis $e_1, \ldots, e_n$ and let E_- be the subspace spanned by those e_i for which $\langle e_i, e_i \rangle = -1$ or 0. Then $Q(v) \leq 0$ for all $v \in E_-$. If F is any positive definite[5] subspace of E,

[5]A real quadratic form Q is *positive definite* if $Q(v) > 0$ for all nonzero v.

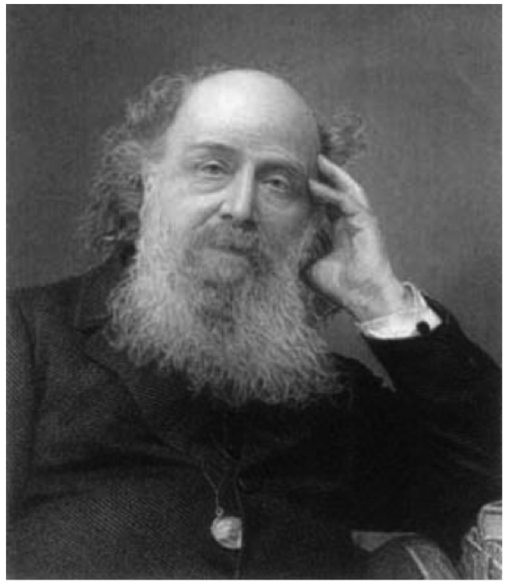

FIGURE 11. James Joseph Sylvester (1814–97) invented the terms 'matrix' and 'discriminant.' Image: MacTutor History of Mathematics Archive

then $F \cap E_- = 0$. According to the forthcoming dimension formula (Theorem 4.6),

$$\dim F + \dim E_- = \dim(F + E_-) + \dim(F \cap E_-).$$

Since $\dim E_- = n - q$, $\dim(F \cap E_-) = 0$, and $\dim(F + E_-) \leq n$, we get that

$$\dim F + n - q \leq n$$

that is, $\dim F \leq q$. Hence

$$\sup\{\dim F \mid F \leq E \text{ positive definite}\} = q,$$

proving that q is independent of the basis chosen. Applying this to the space $(E, -Q)$ shows that p is independent of basis as well. □

Definition 4.4. With the notation of Theorem 4.3, we say that (E, Q) has *Sylvester type* $(-p, q)$.

Corollary 4.5. Two quadratic forms of the same dimension with equal Sylvester type are isometric.

We now state and prove the dimension formula which was crucial in the previous proof.

Theorem 4.6 (Dimension formula)**.** *For U and V subspaces of a finite-dimensional vector space E, we ahve*

$$\dim(U \cap V) + \dim(U + V) = \dim U + \dim V.$$

Proof idea. Apply rank-nullity to the linear map

$$\begin{aligned} f\colon U \oplus V &\longrightarrow E \\ (u, v) &\longmapsto u - v, \end{aligned}$$

observing that $\ker f \cong U \cap V$ and $\operatorname{im} f = U + V$ □

See [Ive92, pp.10–11] for a couple more cute facts about Sylvester types.

Euclidean vector spaces. A *Euclidean vector space* is a finite-dimensional space E equipped with a *positive definite* quadratic form (one for which $Q(v) > 0$ when $v \neq 0$). In such a space, a vector v has *length* $|v| = \sqrt{\langle v, v\rangle}$. Throughtout this section, E stands for a fixed Euclidean vector space.

Theorem 4.7 (Cauchy–Schwarz)**.** *For all* $v, w \in E$,

$$| \langle v, w\rangle | \leq |v||w|$$

and the inequality is strict when v *and* w *are linearly independent.*

Proof. If $w = \lambda v$ for $\lambda \in \mathbb{R}$, then $| \langle v, w\rangle | = |\lambda \langle v, v\rangle | = |\lambda||v|^2 = |v||w|$. If v and w are linearly indpendent, they span a Euclidean plane with Gram matrix $\begin{pmatrix} \langle v, v\rangle & \langle v, w\rangle \\ \langle w, v\rangle & \langle w, w\rangle \end{pmatrix}$. You can read the proof of [Ive92, I.3.5] to see that the sign of the determinant of this matrix is the same as that of any Gram matrix for a form with the same Sylvester type. In particular,

$$\det \begin{pmatrix} \langle v, v\rangle & \langle v, w\rangle \\ \langle w, v\rangle & \langle w, w\rangle \end{pmatrix} > 0$$

so

$$|v|^2|w|^2 - \langle v, w\rangle^2 > 0$$

and the Cauchy-Schwarz inequality follows. ☐

Theorem 4.8 (Triangle inequality)**.** *For all* $v, w \in E$, $|v + w| \leq |v| + |w|$ *and the inequality is strict when* v *and* w *are linearly indpendent.*

Proof. By direct computation,

$$|v + w|^2 = |v|^2 + |w|^2 + 2|v||w| + 2(\langle v, w\rangle - |v||w|).$$

By Theorem 4.7, the term in parentheses is nonnegative (and positive when v, w are linearly independent), and the triangle inequality follows. ☐

Definition 4.9. The *angle* $\angle(v, w)$ between two nonzero vectors v, w in a Euclidean vector space E is

$$\angle(v, w) = \arccos \frac{\langle v, w\rangle}{|v||w|}$$

where $\arccos$ is the inverse to $\cos|_{[0,\pi]}$.

We now turn to the orthogonal group $\mathrm{O}(E)$. If $n \in E$ is a unit normal vector for a hyperlane $H \leq E$ (so $|n| = 1$ and $H \perp n$), then

$$\tau_n(x) = x - 2\langle x, n\rangle n$$

is an orthogonal tranformation of E reflecting across the hyperplane H.

Theorem 4.10. *Suppose* $\dim E = n$. *Then all* $\sigma \in \mathrm{O}(E)$ *can be expressed as the product of at most* n *reflections through hyperplanes.*

Proof. We proceed by induction on $n = \dim E$. The case $n = 1$ is trivial. For the induction step, pick a vector $v \in E$ of unit length. If $\sigma(v) = v$, then σ takes $v^\perp$ to itself and we may use the inductive hypothesis to write σ as a product of at most $n - 1$ reflections. If $\sigma(v) \neq v$, observe that reflection τ along the vector $\sigma(v) - v$ interchanges v and $\sigma(v)$. As such, $\tau \circ \sigma$ fixes the vector v, whence we can write $\tau \circ \sigma$ as a product of at most $n - 1$ reflections. Multiplying on the left by τ expresses σ as a product of at most n reflections. ☐

Recall from Proposition 3.9 that an orthogonal transformation has determinant ± 1. As such, the determinant is a surjective homomorphism $\det\colon \mathrm{O}(E) \to \{\pm 1\}$. The kernel $\mathrm{SO}(E) = \{\sigma \in \mathrm{O}(E) \mid \det E = 1\}$ is thus a normal subgroup of $\mathrm{O}(E)$ of index two called the *special orthogonal group*.

Excercise 4.11. The group $\mathrm{SO}(E)$ acts transitively on $S(E) = \{v \in E \mid |v| = 1\}$ when $\dim E \geq 2$.

Let's now work in the Euclidean plan by assuming $\dim E = 2$. The elements of $\mathrm{SO}(E)$ are then called *rotations*. For e_1, e_2 an orthonormal basis of E and $\sigma \in \mathrm{SO}(E)$ we have

$$\sigma(e_1) = \cos\theta e_1 + \sin\theta e_2$$

for some $\theta \in \mathbb{R}$. Since $\sigma(e_2)$ is orthogonal to $\sigma(e_1)$, we have that the matrix for σ with respect to e_1, e_2 is

$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \qquad \text{or} \qquad \begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix}.$$

Evaluating determinants, we see that the first matrix gives a rotation while the second has determinant -1, so σ has matrix $\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$. By the addition formulæ for trig functions, we get an isomorphism $\mathrm{SO}(E) \cong \mathbb{R}/2\pi\mathbb{Z} \cong S^1 = \{z \in \mathbb{C} \mid |z| = 1\}$. In other words, $\mathrm{SO}(E)$ is a circle.

Now assume that $\dim E = 3$ so we are working in Euclidean 3-space. In this context, we define a *rotation* to be an orthogonal transformation σ of E that fixes a line L and such that $\sigma|_{L^\perp}$ is a (two-dimensional) rotation.

Proposition 4.12 (Euler)**.** If $\dim E = 3$, then every element of $\mathrm{SO}(E)$ is a rotation. If $\sigma \in O(E)$ has $\det\sigma = -1$ then $\sigma = \rho\tau$ where ρ is a rotation with axis L and τ is reflection in the plane $L^\perp$.

Proof. The characteristic polynomial for σ is

$$\chi(t) = \det(t \cdot \mathrm{id} - \sigma).$$

Note that if $\lambda \in \mathbb{R}$ is a real eigenvalue of σ with eigenvector v, then the equation $\sigma(v) = \lambda v$ (along with σ preserving length) implies $\lambda = \pm 1$.

By Proposition 3.9, $\det\sigma = \pm 1$. We also have $\chi(0) = \det(-\sigma) = \mp 1$. Suppose $\det\sigma = 1$ so $\chi(0) = -1$. Since $\chi(t)$ has leading term t^3, we know that $\lim_{t\to\infty}\chi(t) = \infty$. By continuity of χ and the intermediate value theorem, χ has a root on the positive real axis. By the previous paragraph, this root must be 1.

Now suppose $\det\sigma = -1$ so $\chi(0) = 1$. Since $\lim_{t\to-\infty}\chi(t) = -\infty$, we learn that χ has root -1.

In either case, σ has an eigenline L with eigenvalue ± 1. We learn that σ acts on $L^\perp$ with eigenvalue 1 (so that $\det\sigma = \pm 1$), and this is equivalent to Euler's proposition. ☐

The following lemma will allow us to inductively describe the form that elements of $\mathrm{O}(E)$ take according to $\dim E$.

Lemma 4.13. Let $\sigma\colon E \to E$ be any linear endomorphism of a finite-dimensional nontrivial real vector space. There exists a linear subspace $V \leq E$ with $\dim V = 1$ or 2 which is stable under σ.

Read [Ive92, Lemma 4.11] for a cute proof using *complexification*. Recall that every complex linear transformation has an eigenvalue and hence a stable linear subspace. The reader should interpret Lemma 4.13 as a version of this result for real vector spaces (with lines replaced by lines or planes).

Theorem 4.14. *For $\sigma \in \mathrm{O}(E)$ there exists a decompositiion of E into an orthogonal sum of lines and planes stable under σ.*

Proof. By Lemma 4.13, there is a line or plane $V \leq E$ stable under σ. For $v \in V^\perp$ and $w \in V$, choose $w' \in V$ such that $\sigma(w') = w$. Then $\langle\sigma(v), w\rangle = \langle\sigma(v), \sigma(w')\rangle = \langle v, w\rangle = 0$, so $V^\perp$ is stable under σ as well. We may now proceed by induction on $\dim E$. ☐

Parabolic forms. We now drop the definite from our positive definite forms, looking at real quadratic forms Q such that $Q \geq 0$ (but which may have $Q(v) = 0$ for nonzero v). There are called *positive* forms.

Theorem 4.15 (Cauchy-Schwarz)**.** *If Q is positive with polarization $\langle\rangle$, then*

$$\langle v, w\rangle^2 \leq \langle v, v\rangle \langle w, w\rangle .$$

Proof. The same as that of Theorem 4.7. ☐

Corollary 4.16. Let (E, Q) be a positive quadratic form. A vector v is isotropic ($\langle v, v\rangle = 0$) if and only if $v \in E^{\perp}$ (that is, $\langle v, w\rangle = 0$ for all $w \in E$).

We now look at the ortogonal group of such forms.

Theorem 4.17. *Let (E, Q) be a positive quadratic form with $\dim E = m$. If $\sigma \in \mathrm{O}(E)$ and $\sigma|_{E^{\perp}} = \mathrm{id}$, then σ can be written as the product of at most m reflections.*

Proof. Full details are in [Ive92, Theorem 5.3]. This is essentially a more delicate version of Theorem 4.10. ☐

Definition 4.18. A positive quadratic form (E, Q) is called *parabolic* if $\dim E^{\perp} = 1$. In this case, $E^{\perp}$ is called the *isotropic line*.

Observe that $E^{\perp}$ is stable under any $\sigma \in \mathrm{O}(Q)$, and thus when Q is parabolic σ has $E^{\perp}$ as an eigenspace. Define the *multiplier* function

$$\begin{aligned} \mu \colon \mathrm{O}(Q) &\longrightarrow \mathbb{R}^{\times} \\ \sigma &\longmapsto \text{eigenvalue of } \sigma \text{ on } E^{\perp}. \end{aligned}$$

This is a group homomorphism and thus

$$\mathrm{O}_{\infty}(Q) := \{\sigma \in \mathrm{O}(Q) \mid \mu(\sigma) > 0\}$$

is an index 2 subgroup of $\mathrm{O}(Q)$. (Indeed, $\mathrm{O}_{\infty}(Q)$ is the kernel of the composition of μ with $\mathbb{R}^{\times} \to \mathbb{R}^{\times}/\mathbb{R}_{>0}$.)

The antipodal map $a \colon E \to E$, $x \mapsto -x$ commutes with all elements of $\mathrm{O}(E)$ and $\mu(a) = -1$. This gives us a decomposition

$$\mathrm{O}(E) \cong \mathrm{O}_{\infty}(E) \times \mathbb{Z}/2\mathbb{Z}.$$

[Iversen goes on for another fives pages — how much of this do we need right now?]

5. The Lorentz group

As hyperbolic geometers we will be unduly obsessed with $(n+1)$-dimensional real quadratic forms of Sylvester type $(-n, 1)$. The archetype of such a form is

$$x \longmapsto x_{n+1}^2 - \sum_{i=1}^{n} x_i^2$$

on $\mathbb{R}^{n+1}$, but we will find coordinate free descriptions useful as well. The *hyperboloid* or *pseudo-sphere* for such a form (E, Q) is

$$S(E) := \{v \in E \mid Q(v) = 1\}.$$

In coordinates, this is the locus of the equation

$$\sum_{i=1}^{n} x_i^2 = x_{n+1}^2 - 1.$$

The following Cauchy–Schwarz type inequality holds on $S(E)$.

Lemma 5.1. If (E, Q) is an $(n+1)$-dimensional real quadratic form of Sylvester type $(-n, 1)$ and $x, y \in S(E)$, then
$$1 \leq |\langle x, y\rangle|$$
and the inequality is strict when x and y are linearly independent in E.

Proof. If x, y are linearly dependent and both in $S(E)$, then $x = \pm y$ and the inequality is trivial. When x, y are lineaerly independent they span a plane P of Sylvester type $(-1, 1)$, $(-1, 0)$, or $(-2, 0)$ (see Lemma 5.2 below). Since P contains x and y which have norm 1, it must be the case that the Sylvester type is $(-1, 1)$. The ensuing Discriminant Lemma 5.2 implies that
$$\langle x, x\rangle \langle y, y\rangle < \langle u, x\rangle^2$$
from which the Cauchy-Scwartz inequality follows. □

Lemma 5.2 (Discriminant lemma)**.** Let P be a plane in E spanned by x, y, where E is an $(n+1)$-dimensional real quadratic form of Sylvester type $(-n, 1)$. Define the discriminant of P to be
$$\Delta := \langle x, x\rangle \langle y, y\rangle - \langle x, y\rangle^2 .$$
Then the Sylvester type of P is
$$\begin{cases} (-1,1) & \text{if } \Delta < 0, \\ (-1,0) & \text{if } \Delta = 0, \\ (-2,0) & \text{if } \Delta > 0. \end{cases}$$

Proof. We begin by proving that P contains a vector of norm -1. Choose an orthonormal basis for E and use it to create a hyperplane V of Sylvester type $(-n, 0)$. By the dimnension formula Theorem 4.6,
$$\dim P + \dim E = \dim(P \cap E) + \dim(P + E).$$
The left-hand side is $n+2$ and $\dim(P+E) \leq n+1$, so $\dim(P \cap E) \geq 1$. It follows that P contains vectors with negative norm, and thus its Sylvester type is $(-1, 1)$, $(-1, 0)$, or $(-2, 0)$. Again by [Ive92, I.3.5] the determinant of a (real) Gram matrix for a form of Sylvester type $(-p, q)$ has sign $(-1)^p$. This implies our discriminant test. □

We can apply these lemmata to prove that the hyperboloid $S(E)$ always has exactly two connected components. In the following proof, refer to Figure 12 to track what is going on. *Challenge*: Use a computer to draw the same picture one dimension up (and then send said picture to me so I can include it in these notes). Better yet, build an interactive demo that allows the user to vary parameters (such as the choices of x and y in the proof below).

Proposition 5.3. The hyperboloid $S(E)$ has exactly two connected components and $x, y \in S(E)$ are in the same connected component if and only if $\langle x, y\rangle > 0$.

Proof. Fix $y \in S(E)$ and let $V = y^\perp$ be the hyperplane orthogonal to x. Then V has Sylvester type $(-n, 0)$ and separates $S(E)$ into two parts:
$$H = \{x \in S(E) \mid \langle x, y\rangle > 0\} \qquad \text{and} \qquad H^- = \{x \in S(E) \mid \langle x, y\rangle < 0\}.$$

It remains to show that H and H^- are connected. They are clearly homeomorphic via the antipodal map, so it suffices to prove that H is connected. We do so by exhibiting a homeomorphism with the open unit disk
$$D := \{x \in V \mid -\langle x, x\rangle < 1\}.$$
The affine line through $x \in H$ and $-y$ is parameterized by $(x+y)t - y$, $t \in \mathbb{R}$, and its intersection $g(x)$ with V occurs when $\langle (x+y)t - y, y\rangle = 0$, so
$$g(x) = \frac{x - \langle x, y\rangle y}{1 + \langle x, y\rangle}.$$

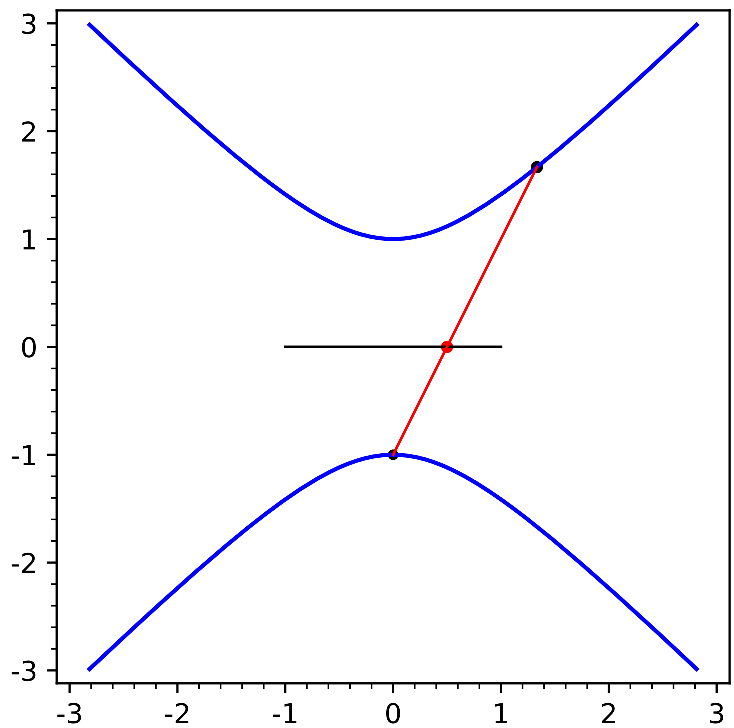

FIGURE 12. This picture gives the content of the proof of Proposition 5.3 when $E = \mathbb{R}^2$ with form $x_2^2 - x_1^2$. The blue curves represent $S(E)$, the black line segment is D, and y has been taken to be $(0,1)$, so H is the upper portion of $S(E)$. The red line segment joins $-y$ to the black dot on H, and the red dot represents the value that g takes.

We leave it to the reader to check (by applying Lemma 5.1) that $g(x) \in D$. One may furthermore check that $g\colon H \to D$ is a homeomorphism by exhibiting its continuous inverse. □

Definition 5.4. A *Lorentz transformation* (or *orthochroneous Lorentz transformation*) of E is any $\sigma \in \mathrm{O}(E)$ which preserves the connected components of $S(E)$. The group of Lorentz transformations is called the *Lorentz group* and is denoted $\mathrm{Lor}(E)$. Lorentz transformations with determinant 1 are called *even* or *special* Lorentz transformations and the group of such maps is denoted $\mathrm{Lor}^+(E)$.

By Proposition 5.3, we know that $\sigma \in \mathrm{O}(E)$ is a Lorentz transformation if and only if

$$\langle x, \sigma(x)\rangle > 0 \qquad \text{for all} x \in S(E).$$

Since the antipodal map is in the center of $\mathrm{O}(E)$ but is not a Lorentz transformation, we get that

$$\mathrm{O}(E) = \mathrm{Lor}(E) \times \mathbb{Z}/2\mathbb{Z}.$$

Suppose that $c \in E$ with $\langle c, c\rangle = -1$. Then $\tau_c(x) = x + 2\langle x, c\rangle c$. We claim that $\tau_c \in \mathrm{Lor}(E)$. Indeed, for $x \in S(E)$,

$$\langle \tau_c(x), x\rangle = \langle x, x\rangle + 2\langle x, c\rangle \cdot \langle c, x\rangle = 1 + 2\langle x, c\rangle^2 > 0$$

so this follows from Proposition 5.3.

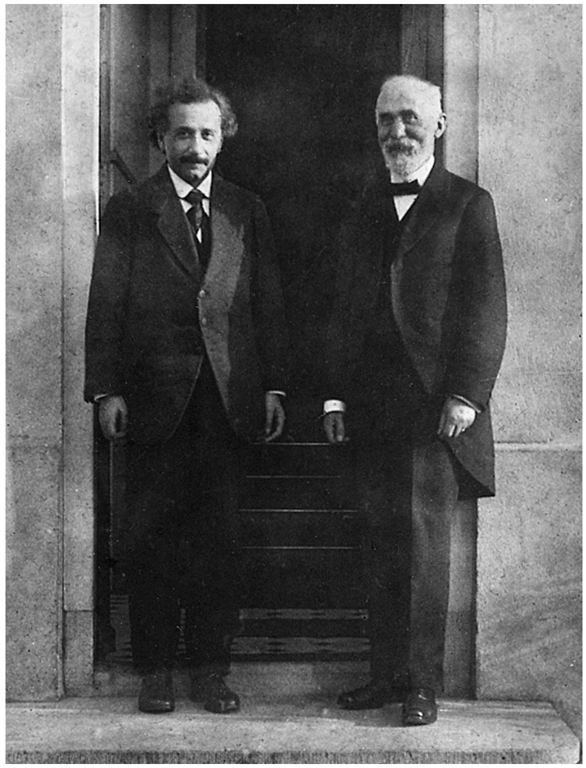

FIGURE 13. Hendrik Lorentz (1853–1928) is the Dutch physicist after whom Lorentz transformations are named. Pictured on the right here in 1921 with Albert Einstein. Image: Wikimedia Commons.

We now study Lorentz transformations in detail when $\dim E = 2$. Pick an orthonormal basis e, f for E with $\langle e, e\rangle = 1$ and $\langle f, f\rangle = -1$. Then a point $xe+tf$ lies in $S(E)$ if and only if $x^2-t^2 = 1$. The component of the hyperbola containing $(1, 0)$ can be parametrized by $s \in \mathbb{R} \mapsto (\cosh s, \sinh s)$. Given $\sigma \in \mathrm{Lor}(E)$ we have $\sigma(e) = e\cosh s + f \sinh s$ for some $s \in \mathbb{R}$. We know that $\langle \sigma(f), \sigma(f)\rangle = -1$ and $\sigma(f) \perp \sigma(e)$, so

$$\sigma(f) = \mp(e \sinh s + f \cosh s).$$

In the first case, the matrix for σ is

$$\begin{pmatrix} \cosh s & -\sinh s \\ \sinh s & -\cosh s \end{pmatrix}$$

with determinant -1 and trace 0, so σ has eigenvalues 1 and -1. It is in fact the case that σ is reflection along a vector of norm -1 (check this!).

In the second case, the matrix for σ is

$$L(s) = \begin{pmatrix} \cosh s & \sinh s \\ \sinh s & \cosh s \end{pmatrix}.$$

Observe that

$$L(s)L(t) = L(s+t)$$

by the following exercise.

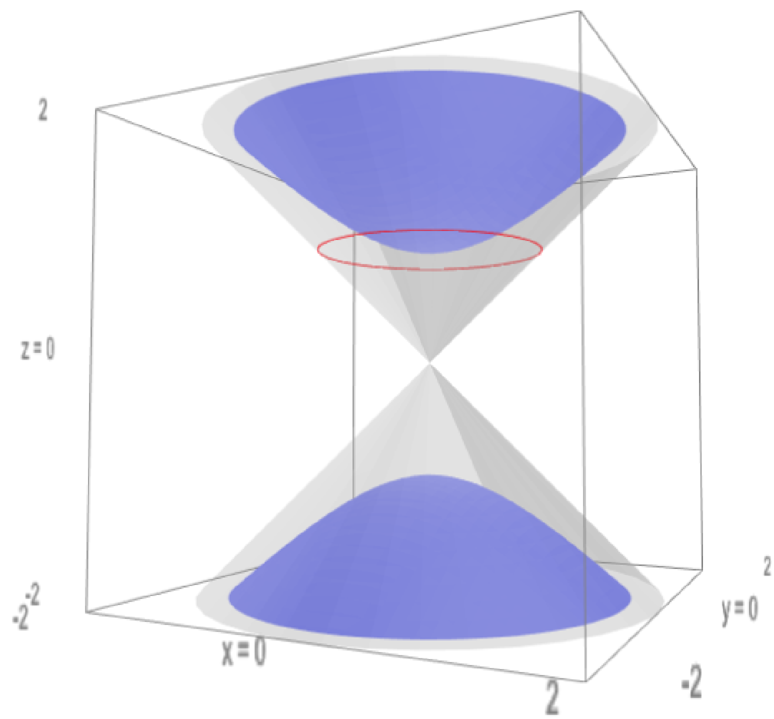

FIGURE 14. The isotropic cone $C(E)$ for $E = \mathbb{R}^3$ with form $z^2 - x^2 - y^2$ is in gray. The blue surface is $S(E)$. The red circle at $z = 1$ can be identified with $PC(E)$ since each line through 0 in $C(E)$ contains a unique point on the circle. In the proof of Proposition 5.7, this circle is instead taken as the boundary of the unit disk in the $z = 0$ plane.

Excercise 5.5. The hyperbolic trigonometric functions satisfy the following angle addition formulæ (which may be checked by direct manipulations of exponential functions):

$$\sinh(s+t) = \sinh s \cosh t + \cosh s \sinh t,$$
$$\cosh(s+t) = \cosh s \cosh t + \sinh s \sinh t.$$

Since $L(s)L(t) = L(s+t)$, we learn that $\mathbb{R} \to \mathrm{Lor}^+(E)$, $s \mapsto L(s)$ is a group isomorphism when $\dim E = 2$.

Theorem 5.6. *Let E be an $(n+1)$-dimensional real vector space equipped with a quadratic form of Sylvester type $(-n, 1)$. Any Lorentz transformation σ of E is the product of at most $n+1$ reflections of the type τ_c with $c \in E$ and $\langle c, c\rangle = -1$.*

Proof. See [Ive92, Theorem I.6.11]. □

For the remainder of this section, we will study the action of $\mathrm{Lor}(E)$ on the *isotropic cone*

$$C(E) = \{v \in E \mid \langle v, v\rangle = 0\}$$

and its projectivization $PC(E)$: The multiplicative group $\mathbb{R}^\times$ acts on $C(E) \smallsetminus \{-\}$ by scalar multiplication, and the orbit space $PC(E) = C(E)/\mathbb{R}^\times$ is the *projective cone* of E. We can think of $PC(E)$ as the set of isotropic lines in E. See Figure 14 for explicit representations of $C(E)$ and $PC(E)$ when $\dim E = 3$.

Proposition 5.7. Let E be an $(n+1)$-dimensional real vector space equipped with a quadratic form of Sylvester type $(-n, 1)$. Then $PC(E)$ is homeomorphic to the sphere $S^{n-1} = \{x \in \mathbb{R}^n \mid |x| = 1\}$.

Proof. Returning to the terminology of the proof of Proposition 5.3, let $S^{n-1} = \partial D$ be the boundary of the unit disk in the hyperplane V perpendicular to some fixed $y \in S(E)$. Define a function

$$\begin{aligned} \phi \colon V &\longrightarrow E \\ z &\longmapsto 2z + y - \langle z, z \rangle \, y. \end{aligned}$$

By direct calculation using $\langle y, z \rangle = 0$, we get

$$\langle \phi(z), \phi(z) \rangle = (1 + \langle z, z \rangle)^2.$$

If $z \in \partial D$, then $\langle z, z \rangle = -1$, whence $\langle \phi(z), \phi(z) \rangle = 0$. As such, ϕ induces a map $\partial D \to C(E)$ which we may compose with the projection $C(E) \to PC(E)$ to get a map $\Phi \colon \partial D \to PC(E)$. The rest of [Ive92, Proposition I.6.12] justifies why Φ is a homeomorphism. □

For the statement of the following corollary, recall that the action of a group G on a set X is *faithful* if $g \cdot x = x$ for all $x \in X$ implies that $g = e$. (That is, only the identity element acts trivially on X.)

Corollary 5.8. Suppose $\dim E \geq 3$. Then
(a) the Lorentz group acts faithfully on $PC(E)$, and
(b) the truncated isotropic cone $C^\times(E) = C(E) \smallsetminus \{0\}$ has two connected components and these are preserved by the $\mathrm{Lor}(E)$.

Proof. See [Ive92, Corollary I.6.14 and Proposition I.6.15]. □

Remark 5.9. In the near future, we will define hyperbolic space $\mathbb{H}^n$ to be one of the two sheets of the hyperboloid $S(E)$. The projective space of E is $P(E) = (E \smallsetminus \{0\})/\mathbb{R}^\times$, the set of lines (through the origin) in E. The projective cone $PC(E)$ is naturally a subspace of $P(E)$. Let $PS(E)$ denote the subset of $P(E)$ consisting of lines in E spanned by vectors of strictly positive norm. The $\partial PS(E) = PC(E)$.

There is a natural homeomorphism $\mathbb{H}^n \to PS(E)$ assigning $x \in \mathbb{H}^n$ to the line spanned by x. Under this identification, it makes sense to write

$$\partial \mathbb{H}^n = PC(E).$$

Remark 5.10. At this point, Iversen continues his first chapter covering Möbius transformations, inversive products of spheres, and the Riemann sphere. We are going to jump ahead to Chapter II next, looking at metric geometry. We will make it all the way to section II.5 on the Klein disk and then circle back to I.7 before studying the Poincaré disk and upper half space.

6. Metric spaces and their isometries

Recall that a *metric space* is a set X equipped with a function $d \colon X \times X \to \mathbb{R}$ satisfying the following properties:

» identity of indiscernibles: $d(x, y) = 0 \iff x = y$,
» symmetry: $d(x, y) = d(y, x)$, and
» triangle inequality: $d(x, z) \leq d(x, y) + d(y, z)$.

An *isometry* between metric spaces X and Y is a surjective distance-preserving function $\sigma \colon X \to Y$.

Excercise 6.1. Prove that every isometry is injective and hence a bijection. Show that the inverse to an isometry is also an isometry. (It follows that the set of self-isometries of a metric space X forms a group $\mathrm{Isom}(X)$ under composition.)

The set of real numbers $\mathbb{R}$ with the metric $d(x, y) = |y - x|$ is a metric space. The following proposition describes the isometries of $\mathbb{R}$.

Proposition 6.2. An isometry $\sigma\colon \mathbb{R} \to \mathbb{R}$ is either a *translation* $x \mapsto x + a$ or a *reflection* $x \mapsto b - x$.

Proof. Since $1 = d(\sigma(0), \sigma(1))$, we know that $\sigma(1) = \sigma(0) \pm 1$. In the $+$ case take τ to be translation by $\sigma(0)$, and in the $-$ case take τ to be the reflection $x \mapsto \sigma(0) - x$. We are going to prove that $\sigma = \tau$. Suppose for contradiction that there is some $c \in \mathbb{R}$ such that $\sigma(c) \neq \tau(c)$. Note that

$$d(\sigma(c), \sigma(0)) = d(c, 0) = d(\tau(c), \tau(0)) = d(\tau(c), \sigma(0))$$

so $\sigma(0)$ is the midpoint of $\sigma(c)$ and $\tau(c)$. Similarly $\sigma(1) = \tau(1)$ is the midpoint of $\sigma(c)$ and $\tau(c)$, but this contradicts $\sigma(0) \neq \sigma(1)$. □

Definition 6.3. Let $J \subseteq \mathbb{R}$ be an interval and X a metric space. A curve $\gamma\colon J \to X$ is a *geodesic curve* if for all $c \in J$ there is an open neighborhood $U \subseteq J$ of c such that $\gamma|_U$ is distance preserving.

Lemma 6.4. Let u be a point of the open interval $J \subseteq \mathbb{R}$ and let $\gamma\colon J \to \mathbb{R}$ be a geodesic curve with $\gamma(u) = 0$. Then $\gamma(t) = \varepsilon(t - u)$ for $\varepsilon = \pm 1$.

Proof. By Proposition 6.2, γ is locally affine on some open $U \subseteq J$ and thus $\gamma'|_U = \pm 1$. Since J is connected, $\gamma' = \pm 1$ on all of J. The rest is calculus. □

7. Euclidean space

Our current goal is to understand the isometries of a Euclidean space E of dimension n. Here the *length* of a vector $v \in E$ is $|v| = \sqrt{\langle v, v\rangle}$ and the *distance* between two points $v, w \in E$ is

$$d(v, w) = |v - w|.$$

The Cauchy-Schwarz Theorem 4.7 implies that E is a metric space.

Proposition 7.1. Every geodesic curve $\gamma\colon \mathbb{R} \to E$ has the form

$$\gamma(t) = et + v$$

where v is some point in E and $e \in E$ has length 1.

Proof. It is easy to check that $t \mapsto et + v$ is a geodesic. For the converse, suppose that γ is a geodesic. For $a \in \mathbb{R}$, choose an open interval $J \subseteq \mathbb{R}$ containing a on which γ preserves distance. By the sharp triangle inequality, any three points $a, b, c \in J$ have $\gamma(a), \gamma(b), \gamma(c)$ on an affine line. It follows that $\gamma(J)$ is contained in an affine line. By Proposition 6.2, we can find e of length 1 such that

$$\gamma(t) = e(t - a) + \gamma(a)$$

on J. As such, γ is a differentiable curve with locally constant derivative. Since $\mathbb{R}$ is connected, γ' is constant on $\mathbb{R}$. The rest is calculus. □

We will now determine the group of isometries of a Euclidean space E, $\mathrm{Isom}(E)$. One example is *reflection* τ in an affine hyperplane $H \subseteq E$. If n is a unit normal vector to H and $u \in H$, then

$$\tau(x) = x - 2\langle x - u, n\rangle n.$$

Lemma 7.2. Suppose $A_1, \ldots, A_p$ and $B_1, \ldots, B_p$ are sequences of points of E such that

$$d(A_i, A_j) = d(B_i, B_j), \qquad 1 \leq i, j \leq p.$$

There there exists $\sigma \in \mathrm{Isom}(E)$ which is a product of at most p reflections such that $\sigma(A_i) = B_i$.

Proof. We proceed by induction on p. If $p = 1$, reflect through the bisecting hyperplane

$$\{x \in E \mid d(x, A_1) = d(x, B_1)\}.$$

For the induction step, let ρ be a product of at most $p - 1$ reflections taking $A_1, \ldots, A_{p-1}$ to $B_1, \ldots, B_{p-1}$. If $\rho(A_p) = B_p$, we are done. If $\rho(A_p) \neq B_p$, observe that

$$d(\rho(A_p), B_i) = d(\rho(A_p), \rho(A_i)) = d(A_p, A_i) = d(B_p, B_i), \qquad 1 \le i \le p-1.$$

This implies that $B_1, \ldots, B_{p-1}$ lie on the hyperplane bisecting $\rho(A_p)$ and B_p. Take τ to be the reflection through this hyperplane. Then $\sigma = \tau\rho$ does the job. □

Definition 7.3. A *simplex* in $E \cong \mathbb{R}^n$ is a sequence $A_0, A_1, \ldots, A_n \in E$ not contained in an affine hyperplane.

Remark 7.4. It might be more standard to call the convex hull of such a collection of points an n-simplex. In our terminology, 0 simplex is a point, the convex hull of a 1-simplex is a line segment, the convex hull of a 2-simplex is a triangle, and the convex hull of a 3-simplex is a tetrahedron.

Lemma 7.5. Let $A_0, \ldots, A_n$ be a simplex of E. If $\alpha, \beta \in \mathrm{Isom}(E)$ and

$$\alpha(A_i) = \beta(A_i), \qquad 0 \le i \le n,$$

then $\alpha = \beta$.

Proof. Set $\sigma = \beta^{-1}\alpha$. It suffices to prove that $\sigma = \mathrm{id}$, so assume for contradiction that there exists $P \in E$ such that $\sigma(P) \neq P$. Since

$$d(\sigma(P), A_i) = d(\sigma(P), \sigma(A_i)) = d(P, A_i), \qquad 0 \le i \le n,$$

we learn that $A_0, \ldots, A_n$ all belong to the affine hyperplane bisecting P and $\sigma(P)$. This contradicts the hypothesis that $A_0, \ldots, A_n$ is a simplex. □

Corollary 7.6. Let σ be an isometry of E that fixes an affine hyperplane H pointwise. Then either σ is a reflection in H or $\sigma = \mathrm{id}$.

Proof. Assume that $\sigma \neq \mathrm{id}$ and pick $P \in E$ with $\sigma(P) \neq P$. By the proof of the lemma, H is the perpendicular bisector of $\sigma(P)$ and P. If τ is reflection in H, then $\tau\sigma$ fixes P and the points of H. By the lemma, $\tau\sigma = \mathrm{id}$, whence $\sigma = \tau$. □

Theorem 7.7. *If E is a Euclidean space of dimension n and $\beta \in \mathrm{Isom}(E)$, then β can be expressed as the product of at most $n + 1$ reflections.*

Proof. Pick a simplex $A_0, \ldots, A_n$ in E. By Lemma 7.2, there exists an isometry σ which is the composite of at most $n + 1$ reflections with $\sigma(A_i) = \beta(A_i)$, $0 \le i \le n$. By Lemma 7.5, $\sigma = \beta$. □

We are now ready to describe $\mathrm{Isom}(E)$ in terms of $\mathrm{O}(E)$ and translations. For a vector $e \in E$, *translation by* e is the isometry t_e given by $t_e(x) = e + x$. The translations form a subgroup $T(E) \le \mathrm{Isom}(E)$ and $T(E) \cong (E, +)$, the additive group underlying E.

Lemma 7.8. Any isometry $\rho \in \mathrm{Isom}(E)$ has a unique expression of the form

$$\rho = t_e\sigma, \qquad e \in E \text{ and } \sigma \in \mathrm{O}(E).$$

The orthogonal translation σ is called the *linearization* of ρ and is denoted $\vec{\rho}$.

Proof. Let $\mathrm{Isom}'(E)$ denote the isometries of E of the form $t_e\sigma$, $e \in E$, $\sigma \in \mathrm{O}(E)$. The reader may check that

$$\sigma t_e \sigma^{-1} = t_{\sigma(e)}.$$

It follows that $\mathrm{Isom}'(E)$ is closed under composition since

$$t_f \sigma t_e \rho = t_f t_{\sigma(e)} \sigma\rho = t_{v+\sigma(e)}(\sigma\rho)$$

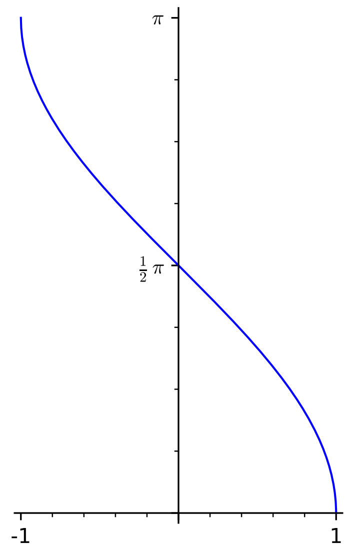

FIGURE 15. The graph of arccos, the inverse of $\cos|_{[0,\pi]}$.

for $f, e \in E$ and $\sigma, \rho \in \mathrm{O}(E)$. By looking at the formula for reflection in an affine hyperplane, we see that all such transformations are in $\mathrm{Isom}'(E)$. It now follows from Theorem 7.7 that $\mathrm{Isom}'(E) = \mathrm{Isom}(E)$, as desired. □

Observe that linearization $\rho \mapsto \vec{\rho}$ is a homomorphism $\mathrm{Isom}(E) \to \mathrm{O}(E)$, and $\vec{\rho} = \mathrm{id}$ if and only if ρ is a translation. This implies that $T(E)$ is the kernel of the linearization map. Since linearization is also surjective, we learn that there is a short exact sequence

$$1 \longrightarrow T(E) \longrightarrow \mathrm{Isom}(E) \longrightarrow \mathrm{O}(E) \longrightarrow 1$$

and that $\mathrm{O}(E) \cong \mathrm{Isom}(E)/T(E)$.

Iversen has a bit more to say about isometries of Euclidean space on pp.63–64 which the interested student may pursue.

8. SPHERICAL GEOMETRY

We will briefly consider the geometry of spheres before moving on to hyperbolic space. Let E be a Euclidean space of dimension $n+1$ with unit sphere $S^n := S(E)$. By the Cauchy–Schwarz Theorem 4.7,

$$\langle P, Q \rangle \in [-1, 1]$$

for $P, Q \in S^n$. As such, we can define the *spherical distance* between $P, Q \in S^n$ to be

$$d(P, Q) = \arccos \langle P, Q \rangle .$$

For the reader's convenience, the graph of arccos is reproduced in Figure 15.

The function d is clearly symmetric. To check identity of indiscernibles, first suppose that P, Q are linearly independent in E. Then the sharp version of Cauchy–Schwarz implies that $\langle P, Q \rangle \in (-1, 1)$, so $d(P, Q) \neq 0$. If P, Q are linearly dependent but distinct, then $P = -Q$, so $\langle P, Q \rangle = -1$ and $d(P, Q) = \pi$. We will prove the triangle inequality after introducing some spherical trigonometry.

Definition 8.1. A *tangent vector* to a point $A \in S^n$ is a vector $T \in E$ with $\langle T, A\rangle = 0$. A tangent vector of length 1 is called a *unit tangent vector*. The space of tangent vectors to A is

$$T_A(S^n) := A^\perp.$$

Lemma 8.2. For $A, B \in S^n$ there exists a unit tangent vector $U \in T_A(S^n)$ such that

$$B = A \cos d(A, B) + U \sin d(A, B).$$

Proof. First suppose that A and B are linearly independent. Take $U \in T_A(S^n) \cap \operatorname{span}\{A, B\}$ of length 1. Then $B \in \operatorname{span}\{A, U\}$, so $B = xA + yU$ for some $x, y \in \mathbb{R}$. Since $\langle A, U\rangle = 0$, we have $x^2 + y^2 = 1$, and hence there is an $s \in [-\pi, \pi]$ such that

$$B = A \cos s + U \sin s.$$

Replacing (s, U) with $(-s, -U)$ if necessary, we may assume that $s \in [0, \pi]$. Then

$$\cos d(A, B) = \langle A, B\rangle = \cos s,$$

so $s = d(A, B)$, as desired.

If A and B are linearly dependent, then $A = \pm B$, so $\sin d(A, B) = 0$ and any unit tangent vector $U \in T_A(S^n)$ will work. □

By a *spherical triangle* $\triangle ABC$ we mean three points $A, B, C \in S^n$ which are linearly independent in E. We define

$$a = d(B, C),$$
$$b = d(A, C),$$
$$c = d(A, B).$$

By Lemma 8.2, we may choose unit tangent vectors $U, V \in T_A(S^n)$ such that

$$B = A \cos c + U \sin c \qquad \text{and} \qquad C = A \cos b + V \sin b.$$

We then define $\angle A = \alpha = \arccos \langle U, V\rangle$, the angle between U and V.

Proposition 8.3 (Spherical Law of Cosines)**.** Using the above notation,

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

Proof. We present the proof when $\dim E = 3$. This adapts to the general case by observing that A, B, C sit on the sphere for their span, which is a 3-dimensional Euclidean space (at least when A, B, C are linearly independent).

By definition, $\cos a = \langle B, C\rangle$. Without loss of generality, we may take $E = \mathbb{R}^3$ and assume that $A = (0, 0, 1)$ and B is in the xz-plane making angle c with the z-axis, so $B = (\sin c, 0, \cos c)$. If C projects to N in the xy-plane, then the angle between N and the x-axis is α, and we conclude that $C = (\sin b \cos \alpha, \sin b \sin \alpha, \cos b)$. We may then compute

$$\langle B, C\rangle = \sin c \sin b \cos \alpha + \cos c \cos b.$$

Equating this and the previous expression for $\langle B, C\rangle$ gives the law of cosines. □

For hints on a coordinate-free proof, see [Ive92, II.3.5].

Theorem 8.4 (Spherical Triangle Inequality)**.** *Given $A, B, C \in S^n$,*

$$d(A, B) \leq d(A, C) + d(C, B)$$

and the inequality is sharp when A, B, C are linearly independent in E.

Proof. We prove the sharp inequality when A, B, C are linearly independent and leave the rest to the reader. In this case, $\cos\alpha < 1$, Proposition 8.3 implies that

$$\cos a < \cos b \cos c + \sin b \sin c = \cos(c - b).$$

For $c - b > 0$, we deduce $a > c - b$, so $c < a + b$. If $c - b < 0$, then $c < a + b$ is trivial since $a > 0$. □

The following description of geodesics in S^n is what one would expect from the "great circle" intuition for such objects. We send the reader to [Ive92, II.3.8] for a full proof.

Theorem 8.5. *Any geodesic curve $\gamma\colon \mathbb{R} \to S^n$ is of the form*

$$\gamma(t) = A \cos t + T \sin t$$

where $A \in S^n$ and T is a unit tangent vector to S^n at A.

We will now describe $\mathrm{Isom}(S^n)$. It is of course that case that any $\sigma \in \mathrm{O}(E)$ induces such an isometry, and we will show that all such isometries are of this form.

Lemma 8.6. Suppose $A_1, \dots, A_p$ and $B_1, \dots, B_p$ are sequences of points in S^n such that $d(A_i, A_j) = d(B_i, B_j)$ for all $1 \leq i, j \leq p$. Then there exists $\sigma \in \mathrm{Isom}(S^n)$ composed of at most p hyperplane reflections with $\sigma(A_i) = B_i$ for $1 \leq i \leq p$.

Proof. Given $A, B \in S^n$, observe that the set of points equidistant from A and B is the intersection of S^n with the linear hyperplane H orthogonal to $A - B$. An orthogonal reflection in H will interchange A and B. The result follows from the argument given in the Euclidean case (Lemma 7.2). □

Theorem 8.7. *We have* $\mathrm{Isom}(S^n) = \mathrm{O}(E)$.

Proof. Suppose $\beta \in \mathrm{Isom}(E)$. Pick a *spherical simplex* in S^n, that is, a sequence $A_0, \dots, A_n$ of points in S^n linearly independent in E. By Lemma 8.6, there is an isometry $\sigma \in \mathrm{O}(E)$ which is the product of at most $n + 1$ reflections in hyperplanes with $\sigma(A_i) = \beta(A_i)$. Now proceed as in the proof of Lemma 7.5. □

We conclude our diversions in spherical geometry by looking at triangles on S^2. Throughout this discussion, we will take it as given that the surface area of a sphere of radius R is $4\pi R^2$, a formula you should have derived in vector calculus. Consider a spherical triangle $\triangle ABC$ with interior angles α at A, β at B, and γ at C. We define the *excess* of this triangle to be $E = \alpha + \beta + \gamma - \pi$.

Theorem 8.8 (Girard)**.** *The area of $\triangle ABC$ is*

$$\mathrm{area}(\triangle ABC) = R^2 \cdot E.$$

You can prove this theorem yourself via the following steps:

(1) Let P_A be the lune[6] built from $\triangle ABC$, the triangle across the BC line segment, and the lune on the opposite side of the sphere. Similarly define P_B and P_C.
(2) Check that $P_A \cup P_B \cup P_C$ is the entire sphere and that the intersection of any two P's is $\triangle ABC \cup$ (a congruent triangle on the other side of the sphere).
(3) Via geometric inclusion-exclusion, deduce that

$$\mathrm{area}(P_A) + \mathrm{area}(P_B) + \mathrm{area}(P_C) = 4\pi R^2 + 4\,\mathrm{area}(\triangle ABC).$$

(4) Argue that $\mathrm{area}(P_A) = 4\alpha R^2$, $\mathrm{area}(P_B) = 4\beta R^2$, and $\mathrm{area}(P_C) = 4\gamma R^2$ to deduce Girard's theorem.

[6] A *lune* is a region between two great circles. Note that every pair of great circles creates four lunes which come in congruent pairs.

FIGURE 16. Albert Girard (1596–1632) contemplating a sphere. Image: http://mathshistory.st-andrews.ac.uk/Biographies/Girard_Albert.html.

We can use Girard's theorem to effectively bound the sum of the interior angles of a spherical triangle. Since the area is always positive, we learn that $E > 0$, and E approaches but never achieves 0 as the triangle becomes very small. In order to place an upper bound on E, we need to ask ourselves some existential questions about triangles: Take a tiny triangle $\triangle ABC$. Look at the exterior of $\triangle ABC$ on the sphere. Is this also a triangle with vertices A, B, C? If you think so, then area is bounded by $4\pi R^2$ and we learn that $E < 4\pi$ (again getting arbitrarily close but not achieving this bound). It is perhaps more conventional to demand that our triangles are always the smaller of the two regions defined by A, B, C. And if A, B, C all lie on a great circle (so that $\triangle ABC$ is one of two halves of the sphere), we should probably view that case as degenerate. With these conventions, we learn that $E < 2\pi$.

The following theorem summarizes the above discussion.

Theorem 8.9. *Let $\triangle ABC$ be a triangle on a sphere of radius R with interior angles α, β, γ. If $\triangle ABC$ is smaller than half the sphere, then*

$$\pi < \alpha + \beta + \gamma < 3\pi.$$

If $\triangle ABC$ is larger than half the sphere, then

$$3\pi < \alpha + \beta + \gamma < 5\pi.$$

9. HYPERBOLIC SPACE

Throughout this section, E is an $(n+1)$-dimensional real vector space equipped with a quadratic form of Sylvester type $(-n, 1)$. According to Proposition 5.3, the hyperboloid $S(E) = \{x \in E \mid \langle x, x\rangle = 1\}$ has two "sheets," that is, connected components homeomorphic to the unit disk in $\mathbb{R}^n$. Let $\mathbb{H}^n$ denote one of these sheets.

In order to put a metric on $\mathbb{H}^n$, note that it follows from Lemma 5.1 and Proposition 5.3 that $\langle P, Q\rangle \geq 1$ for all $P, Q \in \mathbb{H}^n$. As such, we may define the *hyperbolic distance* between P and Q to be

$$d(P, Q) = \operatorname{arccosh} \langle P, Q\rangle .$$

(Recall that $\operatorname{arccosh}$ is the inverse to $\cosh|_{[0,\infty)}$, and $\cosh$ is pictured in Figure 7.)

Symmetry for d is obvious. To check identity of indiscernibles, suppose $P \neq Q$. Then P and Q are linearly independent and thus $\langle P, Q\rangle > 1$, so $d(P,Q) > 0$. We check the triangle inequality in the next theorem.

Theorem 9.1 (Hyperbolic triangle inequality). *For any $A, B, C \in \mathbb{H}^n$,*

$$d(A,B) \leq d(A,C) + d(C,B)$$

and the inequality is strict whenever A, B, C are linearly independent in E.

Proof. Set $a = d(B,C)$, $b = d(C,A)$, and $c = d(A,B)$. Then the Gram matrix for A, B, C is

$$\Delta := \det \begin{pmatrix} 1 & \cosh a & \cosh b \\ \cosh a & 1 & \cosh c \\ \cosh b & \cosh c & 1 \end{pmatrix}.$$

Expansion along the first row gives

$$\begin{aligned}
\Delta &= (1 \cosh^2 c) - \cosh a(\cosh a - \cosh b \cosh c) + \cosh b(\cosh a \cosh c - \cosh b) \\
&= 1 - \cosh^2 a - \cosh^2 b - \cosh^2 c + 2\cosh a \cosh b \cosh c \\
&= (\cosh^2 b - 1)(\cosh^2 c - 1) - (\cosh b \cosh c - \cosh a)^2 - \sinh^2 b \sinh^2 c - (\cosh b \cosh c - \cosh a)^2 \\
&= (\cosh a - \cosh b \cosh c + \sinh b \sinh c)(\cosh b \cosh c + \sinh b \sinh c - \cosh a) \\
&= (\cosh a - \cosh(b-c))(\cosh(c+b) - \cosh a) \\
&= 4\sinh(\frac{1}{2}(a+b+c))\sinh(\frac{1}{2}(a+c-b))\sinh(\frac{1}{2}(a+b-c))\sinh(\frac{1}{2}(c+b-a)).
\end{aligned}$$

Setting $p = \frac{1}{2}(a+B+c)$, we get

$$\Delta = 4\sinh p \sinh(p-a)\sinh(p-b)\sinh(p-c).$$

If A, B, C are linearly independent in E, then they span a subspace of type $(-2,1)$,[7] so $\Delta > 0$.[8] Without loss of generality, suppose $c \geq a, b$. Then

$$p - a = \frac{1}{2}(c-a) + \frac{1}{2}b > 0 \qquad \text{and} \qquad p - b = \frac{1}{2}(c-b) + \frac{1}{2}a > 0.$$

Since $\Delta, p > 0$, we conclude from the factorization of Δ that $p - c > 0$, whence

$$a + b - c = 2(p-c) > 0,$$

which is the strict triangle indequality.

If A, B, C are linearly dependent, then $\Delta = 0$. We leave the rest to the reader. □

Definition 9.2. A *tangent vector* T to a point $A \in \mathbb{H}^n$ is a vector $T \in E$ with $\langle T, A\rangle = 0$. If $\langle T, T\rangle = -1$, then we call T a *unit tangent vector*. The space of tangent vectors to $\mathbb{H}^n$ at A form a hyperplane $T_A(\mathbb{H}^n) = A^\perp$ with Sylvester type $(-n, 0)$.

Lemma 9.3. For all $A, B \in \mathbb{H}^n$, there exists a unit tangent vector $U \in T_A(\mathbb{H}^n)$ such that

$$B = A\cosh d(A,B) + U \sinh d(A,B).$$

[7] Exercise: Check this!

[8] Indeed, the sign of the determinant of the Gram matrix of a real quadratic form of Sylvester type $(-s, r)$ is $(-1)^s$. See [Ive92, I.3.5] for a proof.

Proof. If $A = B$ then $\sinh d(A, B) = 0$ and we can use any unit vector $U \in T_A(\mathbb{H}^n)$. If $A \neq B$, then A and B are linearly independent and span a plane R of Sylvester type $(-1, 1)$ by Lemma 5.2. (We have $\Delta < 0$ because $\langle A, A\rangle = \langle B, B\rangle = 1$ and $\langle A, B\rangle^2 > 1$ by Lemma 5.1.) As such, there is a unit vector $U \in R \cap T_A(\mathbb{H}^2)$ and we may write

$$B = Ax + Uy$$

for some $x, y \in \mathbb{R}$. Since $\langle A, U\rangle = 0$, we get $x^2 - y^2 = 1$, so there is some s such that

$$B = A \cosh s + U \sinh s.$$

Replacing (U, s) with $(-U, -s)$ if necessary we may assume that $s \geq 0$, whence

$$\cosh d(A, B) = \langle A, B\rangle = \cosh s$$

so $s = d(A, B)$ as required. □

The lemma gives us a good guess as to what geodesics in $\mathbb{H}^n$ might be. Indeed, given $A \in \mathbb{H}^n$ and a unit tangent vector $U \in T_A(\mathbb{H}^n)$, we can define a curve

$$\begin{aligned} \gamma \colon \mathbb{R} &\longrightarrow \mathbb{H}^n \\ s &\longmapsto A \cosh s + U \sinh s. \end{aligned}$$

We can then calculate the hyperbolic distance between $\gamma(s)$ and $\gamma(t)$ for $s, t \in \mathbb{R}$: we have

$$\cosh d(\gamma(t), \gamma(s)) = \langle \gamma(t), \gamma(s)\rangle = \cosh s \cosh t - \sinh s \sinh t = \cosh(t - s) = \cosh |t - s|,$$

where the penultimate equality is a standard hyperbolic trig identity. It follow that $d(\gamma(s), \gamma(t)) = |t - s|$, so γ is a geodesic curve.

Note that the image of γ is contained in the plane R spanned by A and U, and Lemma 9.3 implies that the image is exactly $\mathbb{H}^n \cap R$.

Proposition 9.4. Any geodesic curve $\gamma \colon \mathbb{R} \to \mathbb{H}^n$ is of the form

$$\gamma(t) = A \cosh t + U \sinh t$$

where $A \in \mathbb{H}^n$ and U is a unit tangent vector to $\mathbb{H}^n$ at A.

Proof. Suppose that γ is a geodesic and then fix a point $u \in \mathbb{R}$ and pick an open interval $J \subseteq \mathbb{R}$ containing u and such that $\gamma|_J$ is distance preserving. By the sharp triangle inequality (Theorem 9.1), for any three distinct points $r, s, t \in J$, the points $\gamma(r), \gamma(s), \gamma(t)$ are linearly dependent. In particular, if we fix $c, d \in J$ then $\gamma(J)$ is contained in the plane R spanned by $\gamma(c)$ and $\gamma(d)$. Set $A = \gamma(u)$ and pick a unit vector $U \in T_A(\mathbb{H}^n) \cap R$. By Lemma 6.4 and the preceding discussion,

$$\gamma(t) = A \cosh \varepsilon(t - u) + U \sinh \varepsilon(t - u)$$

for $\varepsilon = \pm 1$ and $t \in J$. Thus γ is continuously differentiable on J with $\gamma'(u) = \varepsilon U$. Since $\cosh$ is an even function and $\sinh$ is an odd function, we immediately get

$$\gamma(t) = \gamma(u) \cosh(t - u) + \gamma'(u) \sinh(t - u)$$

for $t \in J$.

Now suppose σ is another geodesic agreeing with γ in a neighborhood of u. It suffices to show that $\sigma = \gamma$ on all of $\mathbb{R}$. To this end, consider the set

$$S = \{t \in \mathbb{R} \mid \sigma(t) = \gamma(t),\ \sigma'(t) = \gamma'(t)\}.$$

Continuity implies that S is closed. To see that S is open, suppose that $t_0 \in S$. Then the above discussion implies that both functions are equal to $t \mapsto \gamma(t_0) \cosh(t - t_0) + \gamma'(t_0) \sinh(t - t_0)$ in an open interval containing t_0. Thus S is open, closed, and nonempty. Connectedness of $\mathbb{R}$ implies that $S = \mathbb{R}$, so $\sigma = \gamma$ everywhere. □

We now know the exact form of every geodesic curve $\gamma\colon \mathbb{R} \to \mathbb{H}^n$, and we have in fact learned that these geodesics are globally — not just locally — distance preserving. The interpretation of the image of γ as $\mathbb{H}^n \cap R$ for $R = \operatorname{span}\{A, U\}$ implies that there is a *unique* geodesic line through any two distinct points of $\mathbb{H}^n$.

We conclude this section by looking at the isometries of $\mathbb{H}^n$. Recall that a Lorentz transformation preserves the sheets of $S(E)$ and thus induces an isometry of $\mathbb{H}^n$. We will show that $\operatorname{Isom}(\mathbb{H}^n) \cong \operatorname{Lor}(E)$. To set terminology, we will call a reflection along a vector with negative norm a *Lorentz reflection*.

Lemma 9.5. Let $A_1, \dots, A_p, B_1, \dots, B_p \in \mathbb{H}^n$ with

$$d(A_i, A_j) = d(B_i, B_j), \qquad 1 \le i, j \le p.$$

Then there exists $\sigma \in \operatorname{Isom}(\mathbb{H}^n)$ composed of at most p Lorentz reflections with $\sigma(A_i) = B_i$.

Proof. Given $A, B \in \mathbb{H}^n$ let

$$\{P \in \mathbb{H}^n \mid d(A,P) = d(B,P)\}$$

denote their perpendicular bisector; this is necessarily the intersection of $\mathbb{H}^n$ and the linear hyperplane orthogonal to the vector $N = A - B$. By Lemma 5.1 and Proposition 5.3, we have

$$\langle N, N\rangle = 2 - 2\langle A, B\rangle < 0.$$

Reflection along N is a Lorentz reflection interchanging A and B. The proof now proceeds as that for Lemma 7.2. ☐

Theorem 9.6. *Every isometry of $\mathbb{H}^n$ is the restriction of a Lorentz transformation of the ambient space E, so* $\operatorname{Isom}(\mathbb{H}^n) \cong \operatorname{Lor}(E)$.

Proof. Let $A_0, \dots, A_n$ be a *hyperbolic simplex*, that is, $n+1$ linearly independent points of E all in $\mathbb{H}^n$. By the previous lemma, there exists an isometry σ which is the composite of at most $n+1$ Lorentz reflections such that $\sigma(A_i) = \beta(A_i)$ for $0 \le i \le n$. The proof now proceeds exactly as that of Lemma 7.5. ☐

10. The Klein and Poincaré disks

We now begin the task of producing alternate models of hyperbolic space with desirable visual and analytic properties. The first of these is the Klein disk.

Let D^n denote the open unit disk in an n-dimensional Euclidean space E. Then the space $E \oplus \mathbb{R}$ carries a natural quadratic form of Sylvester type $(-n, 1)$, namely

$$\langle (A,a), (B,b)\rangle = -\langle A, B\rangle + ab, \qquad A, B \in E,\ a, b \in \mathbb{R}.$$

For $\mathbb{H}^n$ we take the "upper" sheet of the hyperboloid $S(E \oplus \mathbb{R})$ consisting of points (A, a) with $-\langle A, A\rangle + a^2 = 1$ and $a > 0$.

Consider the map $p\colon D^n \to \mathbb{H}^n$ taking a point A to the point which is the intersection of $\mathbb{H}^n$ and the line through $(A, 1)$ and 0. This is pictured for $n = 1$ in Figure 17. The reader may check that a formula for p is given by

$$p(A) = \frac{(A,1)}{\sqrt{1 - \langle A, A\rangle}}.$$

The *Klein disk* is D^n equipped with the metric given by transporting the metric for $\mathbb{H}^n$ along p, that is, for $A, B \in D^n$

$$d(A,B) = \operatorname{arccosh} \frac{1 - \langle A, B\rangle}{\sqrt{(1 - \langle A, A\rangle)(1 - \langle B, B\rangle)}}.$$

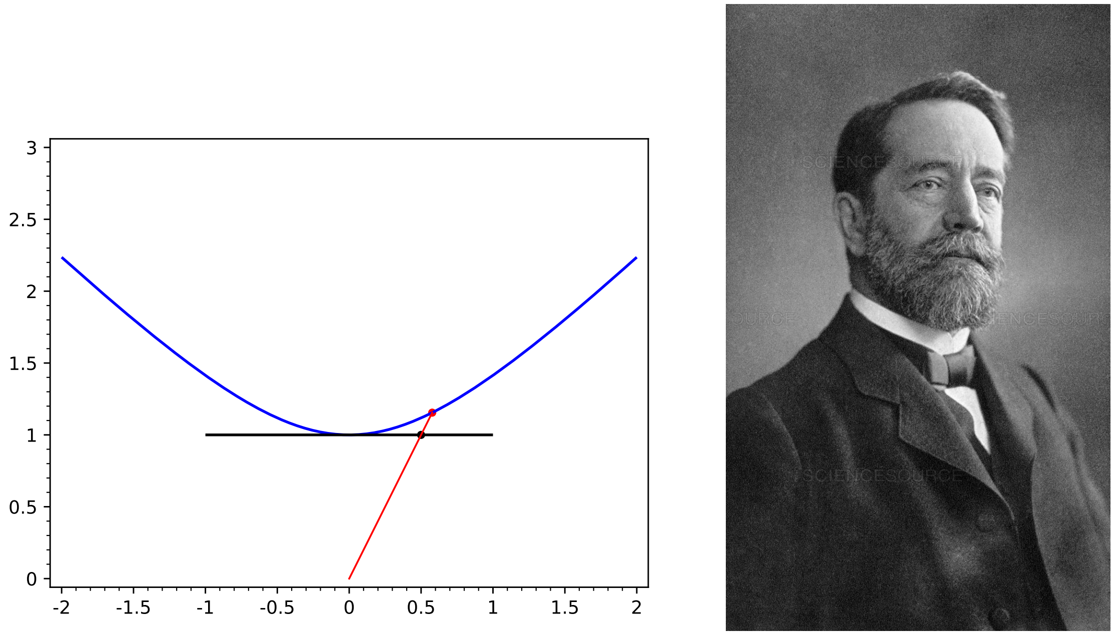

FIGURE 17. Left: A visualization of $p\colon D^n \to \mathbb{H}^n$ when $E = \mathbb{R}$. The disk D^n is drawn at height 1, and the black dot maps to the red dot under p. Right: Felix Klein (1849–1925), the namesake of the Klein disk. Responsible for the Erlangen Program, which classifies geometries by their symmetry groups. Image: Emilio Segrè Visual Archives / American Institute of Physics/Science Source.

Since geodesic lines in $\mathbb{H}^n$ are given by intersections $\mathbb{H}^n \cap R$, R a linear plane in $E \oplus \mathbb{R}$, we see that geodesics in the Klein disk are Euclidean lines in D^n.

For reasons that we will explore later, the Klein disk is not the most desirable model of $\mathbb{H}^n$. A much more attractive model based on the disk is due to Poincaré. Here we again work with $D^n \subseteq E$ and the same $\mathbb{H}^n \subseteq E \oplus \mathbb{R}$, but we parametrize D^n via sterographic projection with center $(0, -1)$. Note that the inverse of this map was previously discussed (under the name g) around Figure 12. A formula for f is given by

$$\begin{aligned} f\colon D^n &\longrightarrow \mathbb{H}^n \\ P &\longmapsto \frac{(2P, 1 + \langle P, P\rangle)}{1 - \langle P, P\rangle}. \end{aligned}$$

Transporting the hyperbolic metric along f results in the *Poincaré disk* with

$$d(P, Q) = \operatorname{arccosh}\left(1 + 2\frac{|P - Q|^2}{(1 - |P|^2)(1 - |Q|^2)}\right)$$

for $P, Q \in D^n$.

In order to say more about the Poincaré disk, we will need to take a diversion into Möbius transformations and the inversive product of spehres, which we undertake in the next section.

11. MÖBIUS TRANSFORMATIONS

Let E be a Euclidean space of dimension n and and let $\mathscr{S}$ be a sphere in E with center c and radius $r > 0$. We can define an involution $\sigma\colon E \smallsetminus \{c\} \to E \smallsetminus \{c\}$ by the formula

$$\sigma(x) = c + r^2 \frac{x - c}{|x - c|^2}.$$

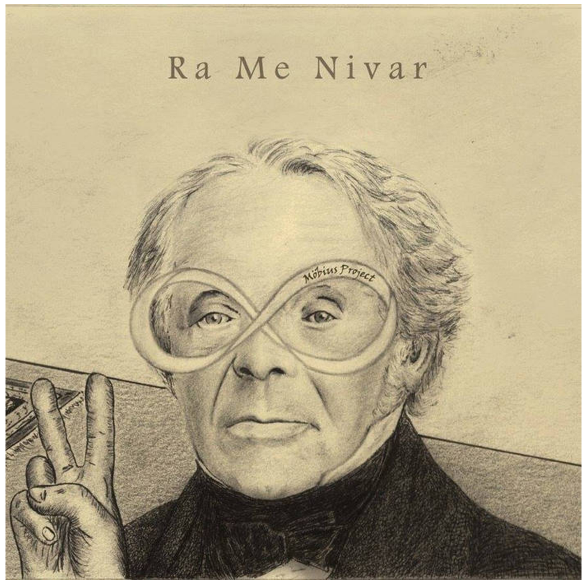

FIGURE 18. Cover art for the Italian prog rock band Möbius Project's 2014 album *Ra Me Nivar*. August Ferdinand Möbius (1790–1868) was a German mathematician and astronomer. In addition to Möbius transformations, he is also the namesake of the Möbius band and Möbius μ-function.

This has the effect of fixing $\mathscr{S}$ pointwise and swapping the interior and exterior of $\mathscr{S}$, so we call σ *inversion* in the sphere $\mathscr{S}$. Note that $\sigma(x)$ lies on the ray originating at c that goes through x.

We can imagine σ sending the center c of $\mathscr{S}$ to a "point at infinity" and also set $\sigma(\infty) = c$. This allows us to extend σ to $\hat{E}$, the *one-point compactification* of E. Consider E as $E \oplus 0 \subseteq E \oplus \mathbb{R}$. Stereographic projection identies $S(E \oplus \mathbb{R})$ with $\hat{E}$, and we give $\hat{E}$ the corresponding topology. The extension $\hat{\sigma}$ of σ to $\hat{E}$ is clearly continuous.

Now consider an affine hyperplane $H \subseteq E$ with unit normal vector n and containing a point $u \in H$. Reflection across H is is given by

$$\begin{aligned} \lambda\colon E &\longrightarrow E \\ x &\longmapsto x - 2\left\langle x-u, n\right\rangle n. \end{aligned}$$

This naturally extends to $\hat{\lambda}\colon \hat{E} \to \hat{E}$ with $\hat{\lambda}(\infty) = \infty$.

By a *sphere* in $\hat{E}$, we will mean either a Euclidean sphere $\mathscr{S}$ in E or $H \cup \{\infty\}$ for H an affine hyperplane in E. We will consider both $\hat{\sigma}$ and $\hat{\lambda}$ to be *inversions in spheres* in $\hat{E}$.

Definition 11.1. The *Möbius group* $\mathrm{M\ddot{o}b}(E)$ is the subgroup of homeomorphisms of $\hat{E}$ generated by inversions in spheres.

Our present goal is to identify $\mathrm{M\ddot{o}b}(E)$ with the Lorentz group of a Minkowski space. To this end, consider the space $E \oplus \mathbb{R}^2$ with form

$$\left\langle (e,a,b),(f,c,d)\right\rangle = -\left\langle e,f\right\rangle + \frac{1}{2}(ad+bc), \qquad e,f \in E,\ a,b,c,d \in \mathbb{R}.$$

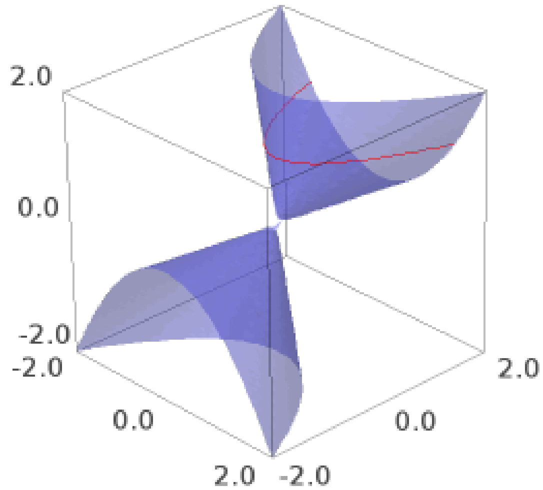

FIGURE 19. This picture shows $C(E \oplus \mathbb{R}^2)$ in blue and the image of ι in red when $E = \mathbb{R}$ with the standard inner product. Note that in this case, the isotropic cone is cut out by the equation $x^2 = yz$ and ι is the curve given by $t \mapsto (t, t^2, 1)$. The fact that the red curve does not make a loop around $C(E \oplus \mathbb{R}^2)$ represents the failure of ι to go through the line spanned by $(0, 1, 0)$.

The form on $\mathbb{R}^2$ given by $\langle (a,b),(c,d)\rangle = \frac{1}{2}(ad+bc)$ has Gram matrix $\begin{pmatrix} 0 & 1/2 \\ 1/2 & 0 \end{pmatrix}$ which has Sylvester type $(-1,1)$. As such, $E \oplus \mathbb{R}^2$ with the above form as Sylvester type $(-n-1,1)$ and is thus a Minkowski space of dimension $n+2$.

Recall that $C(E \oplus \mathbb{R}^2)$ denotes the isotropic cone of vectors with norm 0. We define a map

$$\begin{aligned} \iota\colon E &\longrightarrow C(E \oplus \mathbb{R}^2) \\ e &\longmapsto (e, \langle e,e\rangle, 1). \end{aligned}$$

Indeed,

$$\langle (e, \langle e,e\rangle, 1), (e, \langle e,e\rangle, 1)\rangle = -\langle e,e\rangle + \frac{1}{2}(\langle e,e\rangle + \langle e,e\rangle) = 0,$$

so $f(e)$ is in the isotropic cone. Composing the with natural projection $C(E \oplus \mathbb{R}^2) \to PC(E \oplus \mathbb{R}^2)$ to the projective isotropic cone (the set of lines through 0 in $C(E \oplus \mathbb{R}^2)$) gives a map $\iota\colon E \to PC(E \oplus \mathbb{R}^2)$. The reader may check that this is an injection identifying E with the complement of the point in $PC(E \oplus \mathbb{R}^2)$ corresponding to the line through 0 and $(0,1,0)$. We can extend ι to a bijection $\hat{\iota}\colon \hat{E} \to PC(E \oplus \mathbb{R}^2)$ by defining $\hat{\iota}(\infty)$ to be this line. This is the identification necessary to link $\mathrm{M\ddot{o}b}(E)$ with a Lorentz group.

The group $\mathrm{Lor}(E \oplus \mathbb{R}^2)$ acts on $E \oplus \mathbb{R}^2$ in a manner which takes lines to lines and preserves $\langle\ ,\ \rangle$, and thus induces an action on $PC(E \oplus \mathbb{R}^2) = \partial \mathbb{H}^{n+1}$. Given $\tau \in \mathrm{Lor}(E \oplus \mathbb{R}^2)$, let $PC(\tau)$ denote the corresponding map on $PC(E \oplus \mathbb{R}^2)$. Then $\hat{\tau} = \hat{\iota}^{-1} PC(\tau)\hat{\iota}$ is a map $\hat{E} \to \hat{E}$, as evidenced by the following diagram:

$$\begin{array}{ccc} \hat{E} & \xrightarrow{\hat{\iota}} & PC(E \oplus \mathbb{R}^2) \\ {\scriptstyle \hat{\tau}}\big\downarrow & & \big\downarrow{\scriptstyle PC(\tau)} \\ \hat{E} & \xrightarrow[\hat{\iota}]{} & PC(E \oplus \mathbb{R}^2). \end{array}$$

Theorem 11.2. *The assignment $\tau \mapsto \hat{\tau}$ takes Lorentz transformations of $E \oplus \mathbb{R}^2$ to Möbius transformations of $\hat{E}$ and is an isomorphism of groups* $\mathrm{Lor}(E \oplus \mathbb{R}^2) \cong \mathrm{Möb}(E)$.

Proof. The definition of $\tau \mapsto \hat{\tau}$ immediately implies that this map preserves composition and take the identity to the identity. Since the Lorentz group is generated by Lorentz reflections and the Möbius group is generated by inversions in spheres, it suffices to prove the following three statements: (1) every inversion in a sphere is $\hat{\tau}$ for some $\tau \in \mathrm{Lor}(E \oplus \mathbb{R}^2)$, (2) if τ is a Lorentz reflection, then $\hat{\tau}$ is an inversion in a sphere, and (3) distinct Lorentz transformations induce distinct Möbius transformations.

To prove (1), begin by considering a Euclidean sphere $\mathscr{S} \subseteq E$ with center c and radius $r > 0$. Define the vector $C = (c, \langle c, c \rangle - r^2, 1) \in E \oplus \mathbb{R}^2$. The reader may check that

$$\langle C, C \rangle = -r^2 \qquad \text{and} \qquad 2\langle C, \iota x \rangle = |x - c|^2 - r^2$$

for $x \in E$. With these observations in hand, one can further check that

$$\iota\sigma(x) = \frac{r^2}{|x - c|^2} \tau_C \iota(x)$$

where σ is inversion in $\mathscr{S}$, τ_C is reflection along C, and $x \in E \setminus \{c\}$. This means precisely that $\hat{\tau}_C = \hat{\sigma}$.

Now consider reflection λ in an affine hyperplane $H \subseteq E$ as before. Define the vector $N = (n, 2\langle n, u \rangle, 0)$ for n a unit normal to H and some fixed $u \in H$. The reader may similarly check that $\hat{\tau}_N = \hat{\lambda}$, concluding our proof of (1).

For (2), observe that any vector with negative norm of the form $(e, a, 1) \in E \oplus \mathbb{R}^2$ is of the form C above for a judicious choice of $\mathscr{S}$. Similarly, any vector with negative norm of the form $(e, a, 0)$ is of the form N for some affine hyperplane H.[9] This is enough for (2) since $\tau_c = \tau_{\lambda c}$ for all nonzero scalars λ, and every vector in $E \oplus \mathbb{R}^2$ may be scaled to take the form $(e, a, 1)$ or $(e, a, 0)$.

Finally, (3) follows from Corollary 5.8, which says that the action of the Lorentz group on the projective cone is faithful (if τ acts as the identity, then it is the identity). ☐

Corollary 11.3. A Möbius transformation of a Euclidean space of dimension n can be written as the product of at most $n + 2$ inversions.

Proof. This follows from the above theorem and Theorem 5.6. ☐

We will now analyze the effect of Möbius transformations on disks. Let $N = N(E \oplus \mathbb{R}^2)$ denote the vectors in $E \oplus \mathbb{R}^2$ with norm -1. Given $U \in N$, the set

$$\mathscr{S} = \{x \in \hat{E} \mid \langle \iota x, U \rangle = 0\}$$

is a sphere in $\hat{E}$ with complement $\mathscr{D} \amalg \mathscr{D}^-$ where

$$\mathscr{D} = \{x \in \hat{E} \mid \langle \iota x, U \rangle > 0\} \qquad \text{and} \qquad \mathscr{D}^- = \{x \in \hat{E} \mid \langle \iota x, U \rangle < 0\}.$$

(The reader is encouraged to check that when $U = (e, a, b)$, the sphere $\mathscr{S}$ has center $\frac{1}{b}e$ and radius $|1/b|$.) We call $\mathscr{D}$ and $\mathscr{D}^-$ the *disks bounding* $\mathscr{S}$, and $\mathscr{D}$ is called the *disk with normal vector* U.

[9] Consult [Ive92, p.34] for completely explicit derivations of $\mathscr{S}$ and H.

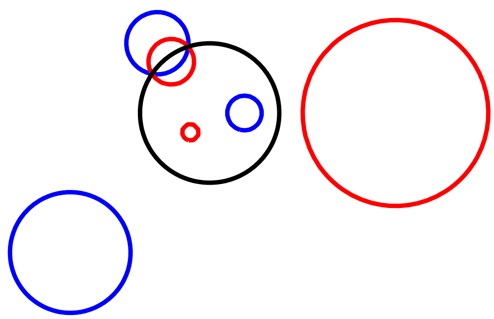

FIGURE 20. This diagram illustrates inversion in the black circle. The red and blue circles are interchanged.

This allows us to deduce the following famous and important corollary. An illustration may be found in Figure 20.

Corollary 11.4. A Möbius transformation $\nu \in \mathrm{Möb}(E)$ takes a sphere $\mathscr{S}$ in $\hat{E}$ to a sphere $\nu(\mathscr{S})$. Inversion in the sphere $\nu(\mathscr{S})$ is given by $\nu\sigma\nu^{-1}$ where σ is inversion in $\mathscr{S}$.

Proof. Write $\overline{\nu} = \iota\nu\iota^{-1}$ for the image of $\nu \in \mathrm{Möb}(E)$ under the inverse of the isomorphism $\mathrm{Lor}(E \oplus \mathbb{R}^2) \cong \mathrm{Möb}(E)$ given by Theorem 11.2. Let U be the normal vector for one of the disks $\mathscr{D}$ bounding $\mathscr{S}$. We claim that $\nu(\mathscr{D})$ is the disk with normal vector $\overline{\nu}(U)$ (which of course implies that $\nu(\mathscr{S})$ is a sphere). Indeed,

$$\begin{aligned}\nu(\mathscr{D}) &= \{\nu(x) \mid \langle \iota x, U\rangle > 0\} \\ &= \{\nu(x) \mid \langle \overline{\nu}\iota x, \overline{\nu}U\rangle > 0\} \quad \text{(since } \overline{\nu} \text{ is an orthogonal transformation)} \\ &= \{\nu(x) \mid \langle \iota\nu x, \overline{\nu}U\rangle > 0\} \quad \text{(definition of } \overline{\nu}\text{)} \\ &= \{y \mid \langle \iota y, \overline{\nu}U\rangle > 0\},\end{aligned}$$

and the final set is the disk with normal vector $\overline{\nu}U$.

For the second part of the corollary, observe that $\overline{\nu} = \tau_U$. Since

$$\tau_{\overline{\nu}(U)} = \overline{\nu}\tau_U\overline{\nu}^{-1}$$

(check this!) we learn that $\nu\sigma\nu^{-1}$ is reflection in the sphere $\nu(\mathscr{S})$. □

Corollary 11.5. The Möbius group $\mathrm{Möb}(E)$ acts transitively on the set of disks in $\hat{E}$.

Proof. By parametrizing the disks by their normal vectors in $N = N(E \oplus \mathbb{R}^2)$, this reduces to checking that $\mathrm{Lor}(E \oplus \mathbb{R}^2)$ acts transitively on N. To see this, note $\mathrm{O}(E \oplus \mathbb{R}^2)$ acts transitively on $N = S_Q(-1)$ by Proposition 3.11. The reader may now check that the stabilizer of any $U \in N$ intersects the Lorentz group trivially, so $\mathrm{Lor}(E \oplus \mathbb{R}^2)$ acts transitively as well. □

Corollary 11.6. Let $\mathscr{S}$ be a sphere in $\hat{E}$ with $\dim E \geq 2$. A Möbius transformation which fixes $\mathscr{S}$ pointwise is either the identity or inversion in $\mathscr{S}$.

Proof. See [Ive92, I.7.12]. □

Definition 11.7. We say that two spheres $\mathscr{S}$ and $\mathscr{T}$ are *orthogonal* if the normal vectors for the bounding disks of $\mathscr{S}$ and $\mathscr{T}$ are orthogonal.

Proposition 11.8. Let σ denote inversion in the sphere $\mathscr{S}$ and let τ denote inversion in the sphere $\mathscr{T}$. If $\mathscr{S} \neq \mathscr{T}$, then the following conditions are equivalent:

(a) $\mathscr{S}$ and $\mathscr{T}$ are orthogonal,
(b) $\sigma\tau = \tau\sigma$,
(c) $\tau(\mathscr{S}) = \mathscr{S}$,
(d) $\sigma(\mathscr{T}) = \mathscr{T}$.

Proof. It follows from Corollary 11.4 that the last three conditions are equivalent. Let H and K be normal vectors for $\mathscr{S}$ and $\mathscr{T}$, respectively. The isomorphism $\mathrm{M\ddot{o}b}(E) \cong \mathrm{Lor}(E \oplus \mathbb{R}^2)$ takes τ to τ_K, Lorentz reflection along K. By the formula

$$\tau_K(H) = H + 2\langle H, K\rangle K$$

we learn that $\tau_K(H) = \pm H$ if and only if $\langle H, K\rangle = 0$. This shows that (a) and (c) are equivalent. □

Definition 11.9. Say that points A and B of $\hat{E}$ are *conjugated* with respect to a sphere $\mathscr{S}$ if $A \neq B$ and inversion in $\mathscr{S}$ interchanges A and B.

Proposition 11.10. If $A, B \in \hat{E}$ are conjugated with respect to a sphere $\mathscr{S}$, then any sphere containing both A and B is orthogonal to $\mathscr{S}$.

Proof. Let $a = \iota A$ and let n denote a normal vector for $\mathscr{S}$. The vector $b = a + 2\langle n, a\rangle n$ generates the isotropic line represented by B (that is, $b \in \hat{\iota} B$). Since $A \notin \mathscr{S}$, $\langle n, a\rangle \neq 0$ and $n \in \mathrm{span}\{a, b\}$. The normal vector for any sphere through A is perpendicular to a and the normal vector for any sphere through B is perpendicula to b. Thus the normal vector for any sphere through A and B is perpendicular to a and b, and thus perpendicular to n (a linear combination of a and b). □

Proposition 11.11. Let $\mathscr{D} \subseteq \hat{E}$ be an open disk. The group $\mathrm{M\ddot{o}b}(\mathscr{D})$ of Möbius transformations of E which take $\mathscr{D}$ to $\mathscr{D}$ is generated by inversions in spheres orthogonal to $\mathscr{S} = \partial\mathscr{D}$.

Proof. Let N denote the normal vector for $\mathscr{D}$. A Möbius transformation σ leaves $\mathscr{D}$ invariant if and only if the corresponding Lorentz transformation $\overline{\sigma}$ fixes N. Thus we may identify $\mathrm{M\ddot{o}b}(\mathscr{D})$ with $\mathrm{Lor}(N^\perp)$. This group is generated by reflections τ_K where $\langle K, K\rangle = -1$ and $\langle K, N\rangle = 0$. □

The following corollary is immediate.

Corollary 11.12. Let H denote a linear hyperplane in E and let $\mathscr{D}$ be one of the half-spaces in E bounded by H. Restriction from $\hat{E}$ to $\hat{H}$ defines an isomorphism $\mathrm{M\ddot{o}b}(\mathscr{D}) \cong \mathrm{M\ddot{o}b}(H)$.

The composite of the inverse of this isomorphism with the inclusion $\mathrm{M\ddot{o}b}(\mathscr{D}) \hookrightarrow \mathrm{M\ddot{o}b}(E)$ is known as *Poincaré extension* $\mathrm{M\ddot{o}b}(H) \to \mathrm{M\ddot{o}b}(E)$.

12. The Poincaré disk and half space

Recall that the Poincaré disk is the open unit disk D^n in an n-dimensional Euclidean space E equipped with the metric

$$d(P, Q) = \mathrm{arccosh}\left(1 + 2\frac{|P-Q|^2}{(1-|P|^2)(1-|Q|^2)}\right).$$

This metric was transported to D^n from $\mathbb{H}^n \subseteq E \oplus \mathbb{R}$ along the map

$$f\colon D^n \longrightarrow \mathbb{H}^n$$
$$P \longmapsto \frac{(2P, 1 + \langle P, P\rangle)}{1 - \langle P, P\rangle}$$

which is stereographic projection from the point $(0, -1)$.

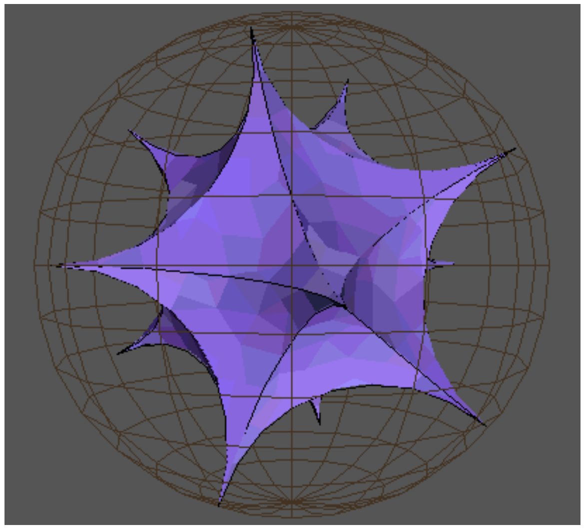

FIGURE 21. An icosahedron in the Poincaré disk D^3. Image: http://home.sandiego.edu/~dhoffoss/rusmp/index.html.

Theorem 12.1. *The group* $\text{Möb}(D^n)$ *of Möbius transformations of* E *that take* D^n *to* D^n *acts as the group of isometries of the Poincaré disk* D^n*. In particular, any isometry of* D^n *extends to a Möbius transformation of* E*.*

Proof. First note that the inverse to f is $f^{-1}\colon \mathbb{H}^n \to D^n$ given by the formula $f^{-1}(A,a) = \frac{A}{1+a}$. Let τ denote reflection along a vector $(N,n) \in E \oplus \mathbb{R}$ of norm -1. The reader may verify that the corresponding transformation of D^n is $f^{-1}\tau f$ given by

$$f^{-1}\tau f(P) = \frac{P + (n + n\langle P,P\rangle - 2\langle P,N\rangle)N}{|N - nP|^2}.$$

When $n = 0$, we have $\langle N,N\rangle = 1$ and $f^{-1}\tau f$ is Euclidean reflection along N. When $n \neq 0$, we have $1 + n^2 = \langle N,N\rangle$ and (by a harder check by the reader) $f^{-1}\tau f$ is inversion in the Euclidean sphere $\mathscr{N}$ with center N/n and radius $1/n$. Since $(1/n)^2 + 1 = \langle N/n, N/n\rangle$, $\mathscr{N}$ is orthogonal to $S^{n-1} = \partial D^n$. Combining this observation with Theorem 9.6, Theorem 5.6, and Proposition 11.11, we are done. □

Proposition 12.2. Every isometry of the Poincaré disk $D^n \subseteq E$ can be written as $\rho\mu$ where $\mu \in \text{O}(E)$ and ρ is a reflection in a sphere orthogonal to $S^{n-1} = \partial D^n$.

The proof given in [Ive92, II.6.5] depends crucially on inversive products of spheres, content we have decided to skip but which may be read in [Ive92, I.8].

The geodesics in the Poincaré disk are arcs of circles which are perpendicular to $S^{n-1} = \partial D^n$ at both intersection points. (Here we include lines as degenerate circles.) This can proven by passing a geodesic in $\mathbb{H}^n$ through $f^{-1}\colon \mathbb{H}^n \to D^n$.

The *Poincaré half space* is constructed as follows. Begin with a Euclidean vector space L of dimension $n-1$ and form the Euclidean space $E = L \oplus \mathbb{R}$. The upper half space E^+ is the set of points $(p,h) \in E$ with $h > 0$. Inversion σ in a sphere of radius $\sqrt{2}$ with center $S = (0,-1)$ maps E^+ onto the unit disk D^n. Furthermore, σ induces stereographic projection of ∂D^n onto $\hat{L} = \partial E^+$. (The reader should pause and convince themselves of the previous two statements.

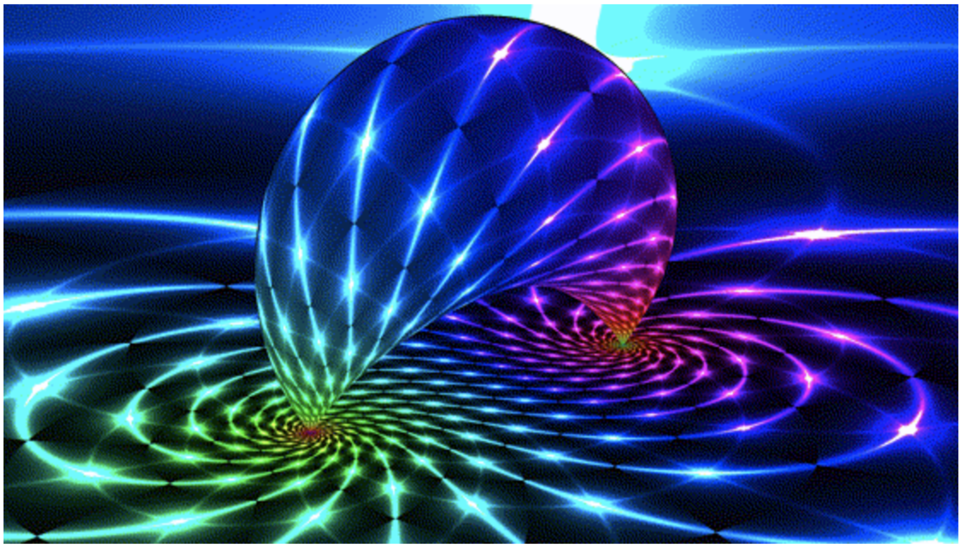

FIGURE 22. A screenshot from http://roice3.org/h3/isometries/ which contains GIF animations of isometries of E^+ for $\dim E = 3$.

Remember that inversions take spheres to spheres, and note that $\sigma(S) = \infty$ while σ pointwise fixes the intersection of its sphere with E^+.) Explicitly,

$$\sigma \colon E^+ \longrightarrow D^n$$
$$(p,h) \longmapsto (0,-1) + 2\frac{(p, 1+h)}{|p|^2 + (1+h)^2}.$$

We can then use σ to transport the metric of D^n onto E^+, resulting in

$$d(P,Q) = \operatorname{arccosh}\left(1 + \frac{|P-Q|^2}{2hk}\right).$$

The upper half space E^+ with the metric d is called *Poincaré half space*.

Theorem 12.3. *The group* $\text{Möb}(E^+)$ *of Möbius transformations of* E *taking* E^+ *to* E^+ *acts as the group of isometries of Poincaré half space. In particular, any isometry of* E^+ *can be extended to a Möbius transformation of* E.

Proof. Conjugation by σ on $\text{Möb}(E)$ takes the subgroup $\text{Möb}(D^n)$ to $\text{Möb}(E^+)$. This theorem then follows from Theorem 12.1. □

Corollary 12.4. The group of isometries of the Poincaré half space E^+ is isomorphic to the full Möbius group of $L = \partial E^+$.

Proof. This follows from the previous theorem and Proposition 11.11. □

We note that the geodesics in E^+ are traced by circles in $\hat{E}$ orthogonal to ∂E^+.

At this point, the reader might enjoy exploring the following virtual reality implementations of hyperbolic 3-space:

http://h3.hypernom.com/
http://www.michaelwoodard.net/hypVR-Ray/

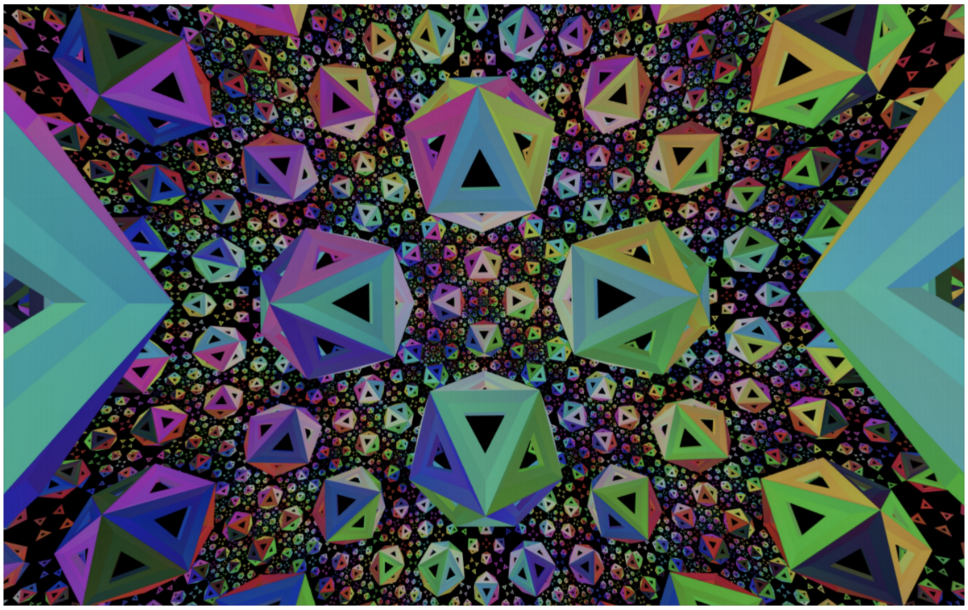

FIGURE 23. A screenshot of h3.hypernom.com.

They are best explored with a VR headset, but should open in any web browser and interact meaningfully with a smart phone (tip: lock screen rotation).

13. THE RIEMANN SPHERE

We need one more auxiliary object before solely focusing on two-dimensional hyperbolic geometry: the Riemann sphere. The *Riemann sphere* $\hat{\mathbb{C}}$ is the one-point compactification of the complex plane. One way to achieve this construction is via projective geometry, as the set of *complex* lines in $\mathbb{C}^2$. This in turn may be realized as the quotient of $\mathbb{C}^2 \setminus \{0\}$ by the action of $\mathbb{C}^\times$ by scalar multiplication. (The orbit of a point (z, w) is $\{(\lambda z, \lambda w) \mid \lambda \in \mathbb{C}^\times\}$, the complex line spanned by (z, w) minus $(0, 0)$.) We will denote the equivalence class of a point $(z, w) \neq (0, 0)$ by $[z : w]$ or $\begin{bmatrix} z \\ w \end{bmatrix}$. This model for $\hat{\mathbb{C}}$ is sometimes denoted $\mathbb{C}P^1$, the *complex projective line*. A point $z \in \mathbb{C} = \hat{\mathbb{C}} \setminus \{\infty\}$ corresponds to $[z : 1]$, while ∞ corresponds to $[1 : 0]$.

The group $\mathrm{GL}_2(\mathbb{C})$ acts on $\hat{\mathbb{C}}$ via the formula

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{bmatrix} z \\ w \end{bmatrix} = \begin{bmatrix} az + bw \\ cz + dw \end{bmatrix}$$

where $a, b, c, d \in \mathbb{C}$, $ad - bc \neq 0$. Note that

$$\begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix} \begin{bmatrix} z \\ w \end{bmatrix} = \begin{bmatrix} \lambda z \\ \lambda w \end{bmatrix} = \begin{bmatrix} z \\ w \end{bmatrix},$$

so scalar matrices in $\mathrm{GL}_2(\mathbb{C})$ acts as the identity. As such, we get an action of

$$\mathrm{PGL}_2(\mathbb{C}) = \mathrm{GL}_2(\mathbb{C})/\mathbb{C}^\times$$

on $\hat{\mathbb{C}}$. (Here we are identifying $\mathbb{C}^\times$ with the normal subgroup of nonzero scalar matrices in $\mathrm{GL}_2(\mathbb{C})$.)

Identifying $\mathbb{C}$ with points of the form $[z:1]$ and ∞ with $[1:0]$, we see that action of $\mathrm{PGL}_2(\mathbb{C})$ on $\hat{\mathbb{C}}$ can also be written as

$$\sigma(z) = \frac{az+b}{cz+d}, \qquad \sigma(-d/c) = \infty, \qquad \sigma(\infty) = \frac{a}{c}$$

for $-d/c \neq z \in \mathbb{C}$ and σ represented by $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$.

Note that for $A \in \mathrm{GL}_2(\mathbb{C})$, the inverse to A is $A^{-1} = \frac{1}{\det A} A^\vee$ where $A^\vee$ is the *cofactor matrix* given by

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^\vee = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

This permits the easy computation of the action of A^{-1} on $\hat{\mathbb{C}}$, especially since we may neglect the determinant factor.

Recall that a group action of G on a set X is called *transitive* if for every $x, y \in X$ there exists $g \in G$ such that $g \cdot x = y$. A more stringent property is the notion of *k-fold transitivity*: for all k-tuples of distinct points $x_1, \dots, x_k \in X$ and $y_1, \dots, y_k \in X$, there exists $g \in G$ such that $g \cdot x_i = y_i$, $1 \leq i \leq k$. In the case of $\mathrm{PGL}_2(\mathbb{C})$ acting on $\hat{\mathbb{C}}$, we actually have a 3-fold (or *triply*) transitive action. The following proposition says this and slightly more.

Proposition 13.1. For any triples A, B, C and P, Q, R of distinct points in $\hat{\mathbb{C}}$, there is a unique $\sigma \in \mathrm{PGL}_2(\mathbb{C})$ transforming the first triple into the second triple.

Proof. Suppose that $A, B, C \in \hat{\mathbb{C}}$ are represented by $a, b, c \in \mathbb{C}^2$. Since a and b are linearly independent, there exist $\lambda, \mu \in \mathbb{C}$ such that $c = \lambda a + \mu b$. Since $A = [\lambda a]$ and $B = [\mu b]$, we may assume without loss generality that $c = a + b$. Similarly, we may represent $P, Q, R \in \hat{\mathbb{C}}$ by $p, q, p+q \in \mathbb{C}^2$. Observe that the linear transformation $\sigma\colon \mathbb{C}^2 \to \mathbb{C}^2$ such that $\sigma(a) = p$ and $\sigma(b) = q$ satisfies $\sigma(a+b) = p+q$. Descending to $\hat{\mathbb{C}}$, we see that the class represented by σ in $\mathrm{PGL}_2(\mathbb{C})$ has the desired properties.

Suppose now that $\tau \in \mathrm{GL}_2(\mathbb{C})$ acts as the identity on $\hat{\mathbb{C}}$. Then all vectors in $\mathbb{C}^2 \setminus \{0\}$ are eigenvectors of τ, whence τ must be a scalar matrix. We conclude that $\mathrm{PGL}_2(\mathbb{C})$ acts faithfully on $\hat{\mathbb{C}}$, and this implies the uniqueness of σ above. $\square$

We have also intensively studied the group of Möbius transformations $\text{Möb}(\mathbb{C})$ which acts on $\hat{\mathbb{C}}$. Call a Möbius transformation *even* if it can be written as the product of an even number of sphere inversions. We write $\text{Möb}^+(E)$ for the subgroup of even Möbius transformations in $\text{Möb}(E)$.

Theorem 13.2. *The group* $\mathrm{GL}_2(\mathbb{C})$ *acts on* $\hat{\mathbb{C}}$ *as even Möbius transformations, inducing an isomorphism*

$$\mathrm{PGL}_2(\mathbb{C}) \cong \text{Möb}^+(\mathbb{C}).$$

Proof. We leave this as a reading exercise. See [Ive92, I.9.5], which depends on [Ive92, I.7.7]. $\square$

We now introduce the cross ratio, an important invariant of four points in $\hat{\mathbb{C}}$ which in some sense measures the failure of quadruple transitivity for the action of $\mathrm{PGL}_2(\mathbb{C})$ on $\hat{\mathbb{C}}$.

Definition 13.3. For points $P, Q, R, S \in \hat{\mathbb{C}}$, at least three of which are distinct, let p, q, r, s be 2×1 complex matrices representing these points. The *cross ratio* of P, Q, R, S is

$$[P, Q; R, S] = [\det(p, r)\det(q, s) : \det(p, s)\det(q, r)] \in \hat{\mathbb{C}}.$$

In a homework problem, you will verify that this value is well-defined and independent of the choice of p, q, r, s.

If $\sigma \in \mathrm{GL}_2(\mathbb{C})$, then
$$\det(\sigma(p), \sigma(q)) = \det(\sigma)\det(p, q).$$
From this, it follows that
$$[\sigma(P), \sigma(Q); \sigma(R), \sigma(S)] = [P, Q; R, S].$$
In other words, the cross ratio is constant on $\mathrm{PGL}_2(\mathbb{C})$-orbits.

Since $[P, Q, R, S]$ is well-defined when at least three of the inputs are distinct, we may fix distinct $P, Q, R \in \hat{\mathbb{C}}$ and then consider the function $\hat{\mathbb{C}} \to \hat{\mathbb{C}}$, $S \mapsto [P, Q, R, S]$ in which S varies and we take the cross ratio.

Proposition 13.4. Let $P, Q, R \in \hat{\mathbb{C}}$ be distinct. Then the transformation
$$\begin{aligned}\hat{\mathbb{C}} &\longrightarrow \hat{\mathbb{C}}\\ S &\longmapsto [P, Q; R, S]\end{aligned}$$
is the action of an element of $\mathrm{PGL}_2(\mathbb{C})$ taking P, Q, R to $\infty, 0, 1$, respectively.

Proof. The reader may check that
$$[P, Q; R, S] = \begin{pmatrix} -q_2 \det(p, r) & q_1 \det(p, r) \\ -p_2 \det(q, r) & p_1 \det(q, r) \end{pmatrix} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix}$$
where the 2×2 matrix has determinant $\det(p, r)\det(q, r)\det(p, q) \neq 0$. The fact that the map takes P, Q, R to $\infty, 0, 1$ is a direct calculation. □

Proposition 13.5. Two 4-tuples of distinct points $P, Q, R, S \in \hat{\mathbb{C}}$ and $X, Y, Z, W \in \hat{\mathbb{C}}$ are in the same $\mathrm{PGL}_2(\mathbb{C})$-orbit if and only if
$$[P, Q; R, S] = [X, Y; Z, W].$$

Proof. By Proposition 13.4, we know that the condition is necessary. Conversely, if the two cross ratios are equal, take $\sigma, \tau \in \mathrm{PGL}_2(\mathbb{C})$ such that $\sigma(P, Q, R) = (\infty, 0, 1)$ and $\tau(X, Y, Z) = (\infty, 0, 1)$. Then
$$[P, Q; R, S] = [\infty, 0; 1, \sigma(S)] = \sigma(S)$$
where the final equality holds because $[\infty, 0; 1, -]$ acts as an element of $\mathrm{PGL}_2(\mathbb{C})$ and takes $\infty, 0, 1$ to $\infty, 0, 1$ (and is thus the identity by Proposition 13.1). Similarly, $[X, Y; Z, W] = \tau(W)$. Thus $\tau^{-1}\sigma$ takes S to W as required. □

Lemma 13.6. Whenever they are defined, the following two identities hold:
$$[P, Q; R, S] = [R, S; P, Q] \qquad \text{and} \qquad [P, Q; R, S] = [Q, P; S, R].$$

Proof. Computation. □

Proposition 13.7. Every element of $\mathrm{PGL}_2(\mathbb{C})$ is conjugate[10] to one of the form
$$\begin{pmatrix} a & 0 \\ 0 & 1 \end{pmatrix} \qquad \text{or} \qquad \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$
for $a \in \mathbb{C}^\times$.

Proof. Let $\sigma \in \mathrm{PGL}_2(\mathbb{C})$ be represented by a matrix $S \in \mathrm{GL}_2(\mathbb{C})$. A nonzero vector $(v, w) \in \mathbb{C}^2$ is an eigenvector for S if and only if $\sigma[v : w] = [v : w]$. By Proposition 13.1, we conclude that for $\sigma \neq \mathrm{id}$, σ has either one or two fixed points in $\hat{\mathbb{C}}$.

[10]Elements g and h of a group G are *conjugate* when there exists $k \in G$ such that $k^{-1}gk = h$.

First suppose that σ has two fixed points, $A, B \in \hat{\mathbb{C}}$. Let $\tau \in \mathrm{PGL}_2(\mathbb{C})$ satisfy $\tau(\infty) = A$, $\tau(0) = B$. Then $\tau^{-1}\sigma\tau$ has fixed points ∞ and 0. Recalling that $\begin{pmatrix} a & b \\ c & d \end{pmatrix} z = \frac{az+b}{cz+d}$, we see that any representing matrix for $\tau^{-1}\sigma\tau$ is diagonal. Multiplying by a complex number, we get a representing matrix of the first type.

Now suppose that σ has only one fixed, $A \in \hat{\mathbb{C}}$. Choose an auxiliary point $C \neq A$ and let τ be the element of $\mathrm{PGL}_2(\mathbb{C})$ satisfying $\tau(\infty) = A$, $\tau(0) = C$, $\tau(1) = \sigma(C)$. Then $\tau^{-1}\sigma\tau$ fixes ∞ and takes 0 to 1. As such, it can be represented by a matrix of the form $\begin{pmatrix} a & 1 \\ 0 & 1 \end{pmatrix}$ for some $a \in \mathbb{C}^\times$. This matrix can only have one eigenvalue, so we must have $a = 1$. □

Given a matrix $S \in \mathrm{GL}_2(\mathbb{C})$, we define

$$\mathrm{tr}^2\, S = \frac{(\mathrm{tr}\, S)^2}{\det S}.$$

For $\lambda \in \mathbb{C}^\times$, we compute

$$\mathrm{tr}^2\, \lambda S = \frac{(\mathrm{tr}\, \lambda S)^2}{\det \lambda S} = \frac{\lambda^2 (\mathrm{tr}\, S)^2}{\lambda^2 \det S} = \mathrm{tr}^2\, S$$

so tr^2 descends to an invariant on $\mathrm{PGL}_2(\mathbb{C})$. It is similarly straightforward to check that tr^2 is constant on conjugacy classes in $\mathrm{PGL}_2(\mathbb{C})$.

Proposition 13.8. Let $\sigma, \tau \in \mathrm{PGL}_2(\mathbb{C}) \smallsetminus \{\mathrm{id}\}$. Then σ and τ are in the same conjugacy class in $\mathrm{PGL}_2(\mathbb{C})$ if and only if $\mathrm{tr}^2\, \sigma = \mathrm{tr}^2\, \tau$.

Proof. The preceding discussion exhibits the necessity of the condition on tr^2. For the converse, note that

$$\mathrm{tr}^2 \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = 4 \qquad \text{and} \qquad \mathrm{tr}^2 \begin{pmatrix} a & 0 \\ 0 & 1 \end{pmatrix} = 2 + a + a^{-1}$$

for $a \in \mathbb{C}^\times$. Since we are assuming $a \neq 1$, we see that these values are distinct, and $2 + a + a^{-1} = 2 + b + b^{-1}$ if and only if $b = a^{\pm 1}$. It remains to show that $\begin{pmatrix} a & 0 \\ 0 & 1 \end{pmatrix}$ and $\begin{pmatrix} a^{-1} & 0 \\ 0 & 1 \end{pmatrix}$ are conjugate, which follows from

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & a \end{pmatrix}$$

and $\begin{pmatrix} a^{-1} & 0 \\ 0 & 1 \end{pmatrix} = a^{-1} \begin{pmatrix} 1 & 0 \\ 0 & a \end{pmatrix}$. □

14. The Poincaré half plane

For the rest of the course, we will focus on hyperbolic geometry in dimension 2. We will do much of this work in the $\mathfrak{sl}_2\mathbb{R}$ model (introduced in the next section), but we pause here to develop the upper half plane model. Let E denote the standard Euclidean space $\mathbb{R}^2$ and write $H^2 = E^+$ for the half-space with second coordinate positive, and identify this space with complex numbers with positive imaginary part. This carries the hyperbolic metric

$$d(z, w) = \operatorname{arccosh}\left(1 + \frac{|w - z|^2}{2\,\mathrm{Im}(z)\,\mathrm{Im}(w)}\right)$$

for $z, w \in \mathbb{C}$. This is equivalent to the following formulæ as the reader may verify:

$$\sinh \frac{1}{2} d(z,w) = \frac{|w-z|}{2\sqrt{\operatorname{Im}(z)\operatorname{Im}(w)}}$$
$$\cosh \frac{1}{2} d(z,w) = \frac{|w-\overline{z}|}{2\sqrt{\operatorname{Im}(z)\operatorname{Im}(w)}}$$
$$\tanh \frac{1}{2} d(z,w) = \frac{|w-z|}{|w-\overline{z}|}.$$

Recall that $\mathrm{GL}_2(\mathbb{R})$ acts on $\mathbb{C} \setminus \mathbb{R}$ via

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

and this has imaginary part

$$\frac{\operatorname{Im} z}{|cz+d|^2} \det \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

As such we can define an action of $\mathrm{GL}_2(\mathbb{R})$ on H^2 via the formula

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} z = \begin{cases} \frac{az+b}{cz+d} & \text{if } ad-bc>0, \\ \frac{a\overline{z}+b}{c\overline{z}+d} & \text{if } ad-bc<0. \end{cases}$$

Proposition 14.1. The above action identifies $\mathrm{PGL}_2(\mathbb{R})$ with $\operatorname{Isom}(H^2)$.

Proof. By Möbius extension, this boils down to $\mathrm{PGL}_2(\mathbb{R})$ being the Möbius group of $\mathbb{R} \subseteq \mathbb{C}$. We leave the details to the reader. $\square$

Proposition 14.2. The geodesics in H^2 are traced by (generalized) circles in $\hat{\mathbb{C}}$ perpendicular to the real axis.

Proof. Note that

$$\cosh(s-t) = \frac{1}{2}(e^{s-t}+e^{t-s}) = \frac{e^{2t}+e^{2s}}{2e^te^s} = 1 + \frac{(e^t-e^s)^2}{2e^te^s} = \cosh d(ie^s, ie^t)$$

for $s, t \in \mathbb{R}$. This implies that $\mathbb{R} \to H^2$, $s \mapsto ie^s$ is a geodesic in H^2, so the positive imaginary axis is a geodesic line. The group $\mathrm{PGL}_2(\mathbb{R})$ acts transitively on circles in $\hat{\mathbb{C}}$ perpendicular to $\mathbb{R}$ (because $\mathrm{PGL}_2(\mathbb{R})$ acts transitvely on $\hat{\mathbb{R}}$). The group $\mathrm{PGL}_2(\mathbb{R})$ is transitive on geodesics in H^2 as well, implying the proposition. $\square$

15. The action of the special linear group on the upper half plane

In Proposition 14.1, we saw that $\mathrm{PGL}_2(\mathbb{R})$ was the group of isometries of the upper half-plane H^2. Every element of $\mathrm{GL}_2(\mathbb{R})$ has determinant in $\mathbb{R}^\times$, and the action of a scalar matrix $\begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix}$ changes this determinant by λ^2. As such, every element of $\mathrm{PGL}_2(\mathbb{R})$ has a representing matrix with determinant ± 1. If the determinant is $+1$, the associated transformation is *orientation-preserving*, and otherwise it is *orientation-reversing*. As such, the group $\mathrm{PSL}_2(\mathbb{R}) = \mathrm{SL}_2(\mathbb{R})/\{\pm I\}$ acts as the group of orientation-preserving isometries of H^2.

Our present goal is to understand some qualitative features of how elements of this group act on H^2. In order to ease notation and nomenclature, we will just discuss the action of $\mathrm{SL}_2(\mathbb{R})$, but the careful reader should remember that $-A$ acts in the same way as A on H^2.

We can understand how $\mathrm{SL}_2(\mathbb{R})$ acts on H^2 by focusing on the fixed points of each transformation. Indeed, for $\sigma \in \mathrm{SL}_2(\mathbb{R})$, $\sigma(z) = z$ if and only if $az + b = (cz + d)z$, *i.e.*,

$$cz^2 + (d - a)z - b = 0. \tag{15.1}$$

By the quadratic formula, this has roots

$$\frac{a - d \pm \sqrt{(d - a)^2 + 4bc}}{2c}.$$

The discriminant $(d - a)^2 + 4bc$ controls the number of real roots and can be expanded as

$$d^2 - 2da + a^2 + 4(ad - 1) = (a + d)^2 - 4 = \operatorname{tr}^2 \sigma - 4.$$

15.1. **Hyperbolic transformations.** When $\sigma \in \mathrm{SL}_2(\mathbb{R})$ has $\operatorname{tr}^2 \sigma - 4 > 0$, (15.1) has two real roots and we call σ a *hyperbolic transformation*. This means that the action of σ on H^2 has no fixed points, but when the action is extended to $H^2 \cup \partial H^2$ there are two distinct fixed points A, B on $\mathbb{R} = \partial H^2$. The geodesic ℓ in H^2 joining A and B is stable under σ, and in fact every (Euclidean) circle through A and B itself by σ. These circles (called *hypercycles* for ℓ) are the orbits of the subgroup of $\mathrm{SL}_2(\mathbb{R})$ consisting of hyerbolic transformations with fixed points A and B.

15.2. **Parabolic transformations.** When $\sigma \in \mathrm{SL}_2(\mathbb{R})$ has $\operatorname{tr}^2 \sigma - 4 = 0$, (15.1) has a single real root and we call σ a *parabolic transformation*. In this case, σ has a single fixed point P on $\mathbb{R} = \partial H^2$ and any circle in $H^2 \cup \partial H^2$ tangent to $\mathbb{R}$ at the point P is stable under the action of σ. Such a circle is a called a *horocycle* with center P, and they form the orbits of the subgroup of $\mathrm{SL}_2(\mathbb{R})$ consisting of parabolic tranformations with fixed point P.

15.3. **Elliptic transformations.** When $\sigma \in \mathrm{SL}_2(\mathbb{R})$ has $\operatorname{tr}^2 \sigma - 4 < 0$, (15.1) has two conjugate complex roots, only one of which is in H^2. In this case, we call σ an *elliptic transformation* and it has a unique fixed point A in H^2 (and none in ∂H^2). The (hyperbolic) circles centered at A are stable under σ. Hyperbolic circles with center A are the orbits of the subgroup of $\mathrm{SL}_2(\mathbb{R})$ consisting of elliptic transformations with fixed point A.

15.4. **Nomenclature.** Why are we calling these transformations hyerbolic, parabolic, and elliptic? The terminology is based on the shape of a plane conic $ax^2 + bxy + cy^2 = 1$, which depends on its discriminant $d = b^2 - 4ac$; it is a hyperbola when $d > 0$, a parabola when $d = 0$, and an ellipse when $d < 0$.

15.5. **Conjugacy classes.** The conjugacy classes in $\mathrm{SL}_2(\mathbb{R})$ are also determined by the sign of $\operatorname{tr}^2 \sigma - 4$. Indeed, we have the following theorem.

Theorem 15.2. *Let $\sigma \in \mathrm{SL}_2(\mathbb{R})$. If $\operatorname{tr}^2 \sigma > 4$, then σ is conjugate to a unique matrix of the form $\begin{pmatrix} \lambda & 0 \\ 0 & 1/\lambda \end{pmatrix}$ with $|\lambda| > 1$. If $\operatorname{tr}^2 \sigma = 4$, then σ is conjugate to exactly one of $\pm I$, $\pm \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$, or $\pm \begin{pmatrix} -1 & 1 \\ 0 & -1 \end{pmatrix}$. If $\operatorname{tr}^2 \sigma < 4$, then σ is conjugate to a unique matrix of the form $\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$ other than $\pm I$.*

A proof based on the action of $\mathrm{SL}_2(\mathbb{R})$ on $\mathbb{R}^2$ can be found in [Con, §3]. Note, though, that an alternate proof can be established via the action of $\mathrm{SL}_2(\mathbb{R})$ on H^2. Indeed, for $\sigma, \tau \in \mathrm{SL}_2(\mathbb{R})$ where σ has (potentially identical) fixed points $A_1, A_2 \in H^2 \cup \partial H^2$, $\tau\sigma\tau^{-1}$ has fixed points $\tau(A_i)$. This and our above analysis shows that the sign of $\operatorname{tr}^2 \sigma - 4$ is preserved by conjugation. To show that any two transformations of the same type (hyperbolic, parabolic, or elliptic) are in the same conjugacy class, one may use Proposition 13.1. The precise conjugacy class representatives may be deduced by specifying "nice" fixed points. For example, the rotation matrices correspond to fixed point i.

15.6. **Construction and visualization.** There are some distinguished families of matrices of each type. Indeed, for $t \in \mathbb{R}$,

$$\begin{pmatrix} e^t & 0 \\ 0 & e^{-t} \end{pmatrix} \quad \text{is hyperbolic with fixed points } 0, \infty,$$
$$\begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix} \quad \text{is parabolic with fixed point } \infty, \text{ and}$$
$$\begin{pmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{pmatrix} \quad \text{is elliptic with fixed point } i.$$

Given (potentially identical) points $A_1, A_2 \in H^2 \cup \partial H^2$, we can conjugate the above matrices by a matrix $C \in \mathrm{SL}_2(\mathbb{R})$ taking their fixed points to A_1, A_2. This results in a family of matrices in $\mathrm{SL}_2(\mathbb{R})$ with the desired fixed points, and we can use this to produce some nice pictures of how elements of $\mathrm{SL}_2(\mathbb{R})$ act on $H^2 \cup \partial H^2$. The reader can find animations produced in this fashion at

http://people.reed.edu/~ormsbyk/341/SL2R.html.

15.7. **Iwasawa decomposition.** We are actually tantalizingly close to deriving the Iwasawa decomposition of $\mathrm{SL}_2(\mathbb{R})$. We send the reader to [Con, Appendix A] for details. (This might make a good final project!)

Consider the following subgroups of $\mathrm{SL}_2(\mathbb{R})$:

$$K = \left\{ \begin{pmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{pmatrix} \,\middle|\, t \in \mathbb{R} \right\},$$
$$A = \left\{ \begin{pmatrix} r & 0 \\ 0 & 1/r \end{pmatrix} \,\middle|\, r > 0 \right\}, \text{ and}$$
$$N = \left\{ \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} \,\middle|\, x \in \mathbb{R} \right\}.$$

Theorem 15.3. *We have a decomposition of* $\mathrm{SL}_2(\mathbb{R})$ *as* $\mathrm{SL}_2(\mathbb{R}) = KAN$*, meaning that every* $g \in \mathrm{SL}_2(\mathbb{R})$ *has a unique representation as* $g = kan$ *with* $k \in K$*,* $a \in A$*, and* $n \in N$*.*

As topological spaces, $K \cong S^1$, $A \cong \mathbb{R}_{>0} \cong \mathbb{R}$, and $N \cong \mathbb{R}$. The function $f\colon K \times A \times N \to \mathrm{SL}_2(\mathbb{R})$ given by $f(k, a, n) = kan$ is a homeomorphism identifying the homeomorphism type of $\mathrm{SL}_2(\mathbb{R})$ as $S^1 \times \mathbb{R}^2$. Since $\mathbb{R}^2 \cong D^2$, the open unit disk, we can (topologically) think of $\mathrm{SL}_2(\mathbb{R})$ as the interior of a solid torus, $S^1 \times D^2$. Beware that $\mathrm{SL}_2(\mathbb{R})$ is not isomorphic to $S^1 \times \mathbb{R}^2$ as a group!

16. Vector calculus on the trace zero model of the hyperbolic plane

Consider the real vector space $\mathfrak{sl}_2(\mathbb{R})$ consisting of 2×2 real matrices with trace 0,

$$\mathfrak{sl}_2(\mathbb{R}) = \left\{ \begin{pmatrix} a & b \\ c & -a \end{pmatrix} \,\middle|\, a, b, c \in \mathbb{R} \right\}.$$

(It turns out that this space is naturally identified with the tangent space at I to $\mathrm{SL}_2(\mathbb{R})$. This makes it the *Lie algebra* of $\mathrm{SL}_2(\mathbb{R})$, whence the lowercase $\mathfrak{mathfrak}$ font.) We compute

$$\begin{pmatrix} a & b \\ c & -a \end{pmatrix} \begin{pmatrix} a & b \\ c & -a \end{pmatrix} = \begin{pmatrix} a^2 + bc & 0 \\ 0 & a^2 + bc \end{pmatrix}$$

so matrices $R \in \mathfrak{sl}_2(\mathbb{R})$ satisfy $R^2 = (-\det R)I$.

We imbue $\mathfrak{sl}_2(\mathbb{R})$ with the symmetric bilinear form

$$\langle R, S \rangle = -\frac{1}{2} \operatorname{tr}(RS).$$

(This bears much similarity to the form you considered in Problem 5 of the first homework assignment, but is slightly different.) We compute

$$\langle R, R\rangle = -\frac{1}{2}\operatorname{tr}(R^2) = -\frac{1}{2}\operatorname{tr}((-\det R)I) = \det R$$

so we may also view $\langle\ ,\ \rangle$ as the polarization of the determinant quadratic form.

We can check the Sylvester type of our inner product by producing an orthonormal basis for $\mathfrak{sl}_2(\mathbb{R})$. The reader is invited to check that

$$\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}, \qquad \begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}, \qquad \begin{pmatrix}0 & 1\\ -1 & 0\end{pmatrix} \tag{16.1}$$

works with determinants -1, -1, 1, respectively. It follows that $\mathfrak{sl}_2(\mathbb{R})$ is a Minkowski space (with Sylvester type $(-2, 1)$).

We now introduce the trilinear form vol on $\mathfrak{sl}_2(\mathbb{R})$, defined by

$$\operatorname{vol}(K, L, M) = -\frac{1}{2}\operatorname{tr}(KLM)$$

for $K, L, M \in \mathfrak{sl}_2(\mathbb{R})$. This form is alternating in the sense that it takes the value 0 when any two of K, L, M are equal. For instance,

$$\operatorname{tr}(KLK) = \operatorname{tr}(LK^2) = -\operatorname{tr}(L\det K) = -\det K\operatorname{tr}L = 0.$$

The form vol is a *volume form* in the sense that it takes the value ± 1 on any orthonormal basis of $\mathfrak{sl}_2(\mathbb{R})$. (This might be a good time to remember the universal property of the determinant of a matrix.) Indeed, we can compute

$$\operatorname{vol}\left(\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}, \begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}, \begin{pmatrix}0 & 1\\ -1 & 0\end{pmatrix}\right) = -\frac{1}{2}\operatorname{tr}\begin{pmatrix}-1 & 0\\ 0 & -1\end{pmatrix} = 1. \tag{16.2}$$

In addition to $\langle\ ,\ \rangle$ and vol, we need the operation $\wedge$ on $\mathfrak{sl}_2(\mathbb{R})$ defined by

$$K \wedge L = \frac{1}{2}(KL - LK)$$

for $K, L \in \mathfrak{sl}_2(\mathbb{R})$. We compute

$$\langle K \wedge L, M\rangle = -\frac{1}{4}\operatorname{tr}(KLM - LKM) = \frac{1}{2}(\operatorname{vol}(K, L, M) - \operatorname{vol}(L, K, M)) = \operatorname{vol}(K, L, M)$$

where the last equality used that vol is alternating and hence $-\operatorname{vol}(L, K, M) = \operatorname{vol}(K, L, M)$.

The following proposition summarizes some relations between these operations.

Proposition 16.3. For all $A, B, C, D, A_i, B_i \in \mathfrak{sl}_2(\mathbb{R})$, $i = 1, 2, 3$, the following identities hold:

(a) $A \wedge (B \wedge C) = \langle A, C\rangle B - \langle A, B\rangle C$,
(b) $\langle A \wedge B, C \wedge D\rangle = \langle A, C\rangle \langle B, D\rangle - \langle A, D\rangle \langle B, C\rangle$, and
(c) $\operatorname{vol}(A_1, A_2, A_3)\operatorname{vol}(B_1, B_2, B_3) = \det(\langle A_i, B_j\rangle)_{i,j}$.

Proof. Identities (a) and (b) are easy computations. For (c), begin by fixing A_1, A_2, A_3. Both sides of the formula are alternating in B_1, B_2, B_3, so we may assume that B_1, B_2, B_3 is a fixed positively oriented orthonormal basis. Now altering A_1, A_2, A_3, we see that both sides are alternating in the A_i, so we can in fact take $A_1 = B_1$, $A_2 = B_2$, and $A_3 = B_3$ to be the same orthonormal basis. We have already computed in (16.2) that the left-hand side is 1 for the basis in (16.1). We also have $\det\operatorname{diag}(-1, -1, 1) = 1$, so the desired equality holds. ☐

17. Pencils of geodesics

The hyperboloid in $\mathfrak{sl}_2(\mathbb{R})$ is the set of trace 0 matrices with determinant 1. If $R = \begin{pmatrix} a & b \\ c & -a \end{pmatrix}$, then $\det R = -a^2 - bc$ and if this quantity is 1, then we know that $c \neq 0$. Thus the sign of c distinguishes between the sheets of the hyperboloid. We let $\mathfrak{h}$ denote the sheet consisting of matrices $R = \begin{pmatrix} a & b \\ c & -a \end{pmatrix}$ with $a^2 + bc = -1$ and $c > 0$. This is the $\mathfrak{sl}_2(\mathbb{R})$ model of the hyperbolic plane.

17.1. **Normal vectors.** Recall that geodesics $\gamma\colon \mathbb{R} \to \mathfrak{h}$ can be parametrized as

$$\gamma(t) = A \cosh t + T \sinh t$$

where $A = \gamma(0)$ and $T = \gamma'(0) \in A^\perp$ with $\langle T, T \rangle = -1$. With this notation, we define the *normal vector* to γ to be

$$N = \gamma'(t) \wedge \gamma(t).$$

More explicitly, we may compute

$$N = (A \sinh t + T \cosh t) \wedge (A \cosh t + T \sinh t) = (T \wedge A)(\cosh^2 t - \sinh^2 t) = T \wedge A,$$

so it is in fact the case that N is independent of t. We have that N is orthogonal to T and A (since $\langle K \wedge L, M \rangle = \operatorname{vol} K, L, M$ and vol is alternating), and the norm of N is

$$\langle T \wedge A, T \wedge A \rangle = \langle T, T \rangle \langle A, A \rangle - \langle A, T \rangle \langle A, T \rangle = -1.$$

We can recover $\gamma'(t)$ from N by the formula

$$\gamma'(t) = \gamma(t) \wedge N.$$

(This follows from the definition of N and Proposition 16.3(a).) Note that N depends on the orientation of γ, so the normal vector of an unoriented geodesic curve is only determined up to sign.

We can now verify that $\mathfrak{h}$ satisfies the reflection axiom!

Theorem 17.1 (Reflection Axiom)**.** *The complement of a geodesic ℓ in $\mathfrak{h}$ has two connected components. Reflection $\tau = \tau_N$ along a normal vector N for ℓ will fix ℓ pointwise and interchange the two connected components of $\mathfrak{h} \setminus \ell$.*

Proof. Let N be a normal vector for ℓ. The function $\mathfrak{h} \to \mathbb{R},\ X \mapsto \langle X, N \rangle$ is zero for $X \in \ell$ and nonzero on $\mathfrak{h} \setminus \ell$. Thus the sets U and V where the function is positive and negative, respectively, separate $\mathfrak{h} \setminus \ell$. If $\tau = \tau_N$, then

$$\tau(X) = X + 2 \langle X, N \rangle N$$

from which it follows that $\langle X, N \rangle = - \langle \tau(X), N \rangle$. This means that U and V are interchanged by τ.

It remains to show that U is connected. Consider two distinct points $A, B \in U$. Pick a tangent vector T to $\mathfrak{h}$ at A such that

$$B = A \cosh d + T \sinh d$$

where $d = d(A, B)$. Let $\gamma(t) = A \cosh t + T \sinh t$, $t \in \mathbb{R}$. Then

$$\langle \gamma(t), N \rangle = \cosh t (\langle A, N \rangle + \langle T, N \rangle \tanh t).$$

If $\langle T, N \rangle \geq 0$, we get that $\gamma(t) \in U$ for all $t \in \mathbb{R}$. If $\langle T, N \rangle < 0$, we can use properties of $\tanh$ to find $r \in \mathbb{R}$ such that $\langle \gamma(t), N \rangle < 0$ for all $t > r$ and $\langle \gamma(t), N \rangle < 0$ for all $t < r$. As such, $\gamma(-\infty, r) \subseteq U$, $\gamma(r, \infty) \subseteq V$, and the geodesic arc $[A, B]$ is a subset of U. This shows that U is path connected and hence connected. ∎

If ℓ is an oriented geodesic curve, then $\mathfrak{h} \setminus \ell$ has two connected components and we can distinguish these *sides* of ℓ by its normal vector N: a geodesic curve through $A \in \ell$ with tangent vector N at A passes from the *negative side* of ℓ to the *positive side* of ℓ.

17.2. **Angles.** Recall that the tangent space of $\mathfrak{h}$ at a point $A \in \mathfrak{h}$ is $T_A(\mathfrak{h}) = A^\perp$, the vectors in $\mathfrak{sl}_2(\mathbb{R})$ orthogonal to A. Call a pair of tangent vectors X, Y at A *positively oriented* if $\mathrm{vol}(A, X, Y) > 0$. Given a unit tangent vector $S \in T_A(\mathfrak{h})$, one can check that $S, S \wedge A$ forms a positively oriented orthonormal basis for $T_A(\mathfrak{h})$.

Given two unit tangent vectors $S, T \in T_A(\mathfrak{h})$, the *oriented angle* $\angle_{\mathrm{or}}(S,T) \in \mathbb{R}/2\pi\mathbb{Z}$ between S and T is defined by

$$T = S \cos \angle_{\mathrm{or}}(S,T) + S \wedge A \angle_{\mathrm{or}}(S,T).$$

17.3. **Intersecting geodesics.** Suppose that ℓ and m are oriented geodesics in $\mathfrak{h}$ passing through a common point A. Set U equal to the tangent vector to ℓ at A and set V equal to the tangent vector to m at A. The *directional angle* from ℓ to m is defined to be $\alpha = \angle_{\mathrm{or}}(V, -U)$. In this case, $-U = V \cos\alpha + V \wedge A \sin\alpha$, and we can compute $\langle U, V\rangle = \cos\alpha$ and $U \wedge V = A \sin\alpha$. For the normal vectors $H = U \wedge A$ and $K = V \wedge A$, we compute

$$\langle H, K\rangle = \cos\alpha, \qquad H \wedge K = A \sin\alpha.$$

In particular, $|\langle H, K\rangle| < 1$. It turns out that the converse is true as well:

Proposition 17.2. Two distinct geodesics ℓ and m in $\mathfrak{h}$ intersect if and only if their normal vectors H and K satisfy $|\langle H, K\rangle| < 1$.

Proof. We have already observed one direction. For the other, suppose that $|\langle H, K\rangle| < 1$. The discriminant for the plane E spanned by H and K is

$$\Delta = \langle H, H\rangle \langle K, K\rangle - \langle H, K\rangle^2 = 1 - \langle H, K\rangle^2 > 0.$$

By the Discriminant Lemma 5.2, E has Sylvester type $(-2, 0)$. Thus $E^\perp$ has type $(0, 1)$ and intersects $\mathfrak{h}$ nontrivially. In fact, the point A which is the intersection of E and $\mathfrak{h}$ is the intersection of ℓ and m. ☐

We say that geodesics ℓ and m are *perpendicular* in $\mathfrak{h}$ if they meet at a point A with angle $\pi/2$.

Corollary 17.3. Geodesics ℓ and m in $\mathfrak{h}$ are perpendicular if and only if their normal vectors are orthogonal.

17.4. **Geodesics with a common perpendicular geodesic.** Suppose ℓ and m are geodesics both perpendicular to another geodesic n. Let A be the point at which ℓ and n intersect, and let B be the point at which m and n intersect. Let N be the normal vector for n corresponding to the orientation of n from A towards B. Use the normal vector $H = A \wedge N$ for ℓ and the normal vector $K = -B \wedge N$ for m. We find that

$$\langle H, K\rangle = \langle A, N\rangle \langle N, B\rangle - \langle A, B\rangle \langle N, N\rangle = \langle A, B\rangle.$$

As such,

$$\langle H, K\rangle = \cosh d(A, B).$$

We also want to show that $\mathrm{vol}\, H, N, K = -\sinh d(A, B)$. To prove this, observe that $N \wedge K = B$ since $K = N \wedge B$. Also $\mathrm{vol}(H, N, K) = \langle H, B\rangle$, so

$$B = A \cosh d(A, B) + H \sinh d(A, B).$$

This gives $\langle H, B\rangle = \langle H, H\rangle \sinh d(A, B) = -\sinh d(A, B)$ as required.

Theorem 17.4. *Two distinct geodesics ℓ and m are perpendicular to the same geodesic if and only if their normal vectors H and K satisfy $|\langle H, K\rangle| > 1$. The common perpendicular geodesic is unique whenever it exists.*

Proof. See [Ive92, III.2.9]. The unique geodesic has normal vector the unique vector of norm -1 in $\operatorname{span}\{H, K\}^\perp$. ☐

17.5. **Lines with a common end.** A geodesic ℓ in $\mathfrak{h}$ spans a plane of type $(-1, 1)$ in $\mathfrak{sl}_2(\mathbb{R})$. The two isotropic lines in this plane are called the *ends* of ℓ. A normal vector H for ℓ is orthogonal to the two ends of ℓ.

Proposition 17.5. Two distinct geodesics ℓ and m in $\mathfrak{h}$ have a common end if and only if their normal vectors H and K satisfy $|\langle H, K\rangle| = 1$.

Proof. See [Ive92, III.2.10]. The common end is $\operatorname{span}\{H, K\}^\perp$. ☐

17.6. **Pencils of geodesics.** A *pencil* is a set $\mathscr{P}$ of geodesics in $\mathfrak{h}$ of the form

$$\mathscr{P} = \{\ell \mid \ell \text{ a geodesic with normal vector in } P\}$$

where P is a fixed linear plane in $\mathfrak{sl}_2(\mathbb{R})$. Note that P is the span of its vectors of norm -1, so P is the span of the normal vectors of geodesics in the pencil it determines. The plane P has Sylvester type $(-2, 0)$, $(-1, 1)$, or $(-1, 0)$, and we will examine the nature of $\mathscr{P}$ according to this type.

If P has type $(-2, 0)$, then the line $P^\perp$ has type $(0, 1)$ and intersects $\mathfrak{h}$ in a point A. The pencil $\mathscr{P}$ is the set of geodesics through A.

If P has type $(-1, 1$, then $P^\perp$ has type $(0, -1)$ and is the span of a normal vector of a geodesic ℓ. The pencil $\mathscr{P}$ is the set of all geodesics perpendicular to ℓ.

If P has type $(-1, 0)$, then $P^\perp$ is isotropic and $\mathscr{P}$ is the set of geodesics with end $P^\perp$.

The following statements now have straightforward verifications. See [Ive92, p.98] for details.

Proposition 17.6. (a) Through two distinct points $A, B \in \mathfrak{h}$, there is a unique geodesic ℓ.
(b) Given a point $A \in \mathfrak{h}$ and a geodesic ℓ, there is a unique geodesic passing through A which is perpendicular to ℓ.
(c) Given a point $A \in \mathfrak{h}$ and an end S, there is a unique geodesic through A with end S.
(d) Given an end S and a geodesic ℓ with ends different from S, there is a unique geodesic perpendicular to ℓ with end S.
(e) There is a unique geodesic with given ends R and S.

18. Classification of isometries

By Theorem 5.6, an isometry of $\mathfrak{h}$ is the product of at most three reflections in geodesics in $\mathfrak{h}$. In particular, every even reflection is a product $\alpha\beta$ of reflections in geodesics a and b. The relative position of a and b (in terms of pencils) will give finer structure on $\operatorname{Isom}^+(\mathfrak{h})$. We will see that odd isometries are easier: they are all *glide reflections*, that is, the composite of a translation along a geodesics and a reflection in the same geodesic.

Theorem 18.1. *Let α, β, γ be reflections in three geodesics a, b, c, respectively, belonging to the same pencil $\mathscr{P}$. Then the product $\alpha\beta\gamma$ is a reflection in a geodesic from $\mathscr{P}$.*

Proof. See [Ive92, p.99]. The argument hinges on considering how α, β, γ act on a nonzero vector in $P^\perp$ where P is the plane in $\mathfrak{sl}_2(\mathbb{R})$ associated with $\mathscr{P}$. ☐

For a triangle $\triangle ABC$ in $\mathfrak{h}$, let m_a, m_b, and m_c be the perpendicular bisectors of BC, AC, and AB, respectively.

Corollary 18.2. The perpendicular bisectors m_a, m_b, m_c for $\triangle ABC$ lie in a pencil.

Proof. Let $\mathscr{P}$ be the pencil containing m_a and m_b and let $\ell \in \mathscr{P}$ be the geodesic through A. Let μ_a, μ_b, and α denote reflection in m_a, m_b, and ℓ, respectively. By Theorem 18.1, $\mu_a\mu_b\alpha$ is a reflection in a geodesic $m \in \mathscr{P}$. But $\mu_a\mu_b\alpha(A) = \mu_a\mu_b(A) = \mu_a(C) = B$, so $m = m_c$, implying that $m_c \in \mathscr{P}$. $\square$

18.1. **Horolations.** Horolations are a form of even isometry that are also referred to as limit rotations or parabolic transformations.

Definition 18.3. Let S be an isotropic line in $\mathfrak{sl}_2(\mathbb{R})$. A *horolation* with center S is an isometry of the form $\alpha\beta$ where α and β are reflections in geodesics a and b with end S.

Note that the set of horolations with center S forms a group. Indeed, the tricky part is closure under multiplication, but this is a corollary of Theorem 18.1. More is true:

Corollary 18.4. The group of horolations with center S is abelian.

Proof. Suppose ρ, σ, τ are reflections in geodesics in $\mathfrak{h}$ with end S. Then $\rho\sigma\tau = (\rho\sigma\tau)^{-1} = \tau^{-1}\sigma^{-1}\rho^{-1} = \tau\sigma\rho$. For four such reflections, we get

$$(\alpha\beta)(\gamma\delta) = (\alpha\beta\gamma)\delta = (\gamma\beta\alpha)\delta = \gamma(\beta\alpha\delta) = \gamma(\delta\alpha\beta) = (\gamma\delta)(\alpha\beta)$$

as desired. $\square$

The reader should consult [Ive92, p.101] for further details on horolations. It is actually the case that the horolations with center S are isomorphic (as a group) to $(\mathbb{R}, +)$. We also have complete control over the orbits of this group acting on $\mathfrak{h}$.

18.2. **Rotations.** Rotations are another class of even isometries of $\mathfrak{h}$ that we have previously called elliptic transformations.

Definition 18.5. A *rotation* around a point A of $\mathfrak{h}$ is an isometry of the form $\alpha\beta$ where α and β are reflections in geodesics a and b passing through A.

Rotations around $A \in \mathfrak{h}$ formm a group isomorphic to $\mathbb{R}/2\pi\mathbb{Z}$. The orbits of this group are the circles centered at A.

18.3. **Translations.** Translations are the class of even isometries of $\mathfrak{h}$ that we previously called hyperbolic transformations.

Definition 18.6. Let k denote a geodesic in $\mathfrak{h}$. A *translation* along k is an isometry of $\mathfrak{h}$ of the form $\alpha\beta$ where α and β are reflections in geodesics a and b perpendicular to k.

Translations along k form a group isomorphic to $\mathbb{R}$. The reader should take a moment to think through the orbits of this group acting on $\mathfrak{h}$.

18.4. **Odd isometries.** The odd isometries of $\mathfrak{h}$ are those which are the composition of an odd number of reflections. They all have a fairly simple form:

Proposition 18.7. Every odd isometry of $\mathfrak{h}$ is a *glide reflection* of the form $\tau\kappa$ where τ is a translation along a geodesic k and κ is reflection in k.

Proof. Let ϕ be an odd isometry of $\mathfrak{h}$. Fix $A \in \mathfrak{h}$. If $A = \phi(A)$, then ϕ is a reflection in a geodesic through A. If $A \neq \phi(A)$, let m denote the geodesic through A and $\phi(A)$ and let ℓ denote the perpendicular bisector for A and $\phi(A)$. Let M denote the midpoint between A and $\phi(A)$, which is also the intersection of ℓ and m. Let μ and λ denote reflection through m and ℓ, respectively. Then $\lambda\phi$ is even and fixes the point A, so $\lambda\phi$ is a rotation around A. By Theorem 18.1, $\lambda\phi = \mu\nu$ where ν is a reflection in a geodesic n through A. Let k be the geodesic through M perpendicular to n and let κ denote reflection in k. Then

$$\phi = \lambda(\mu\nu) = (\lambda\mu)(\nu\kappa)\kappa.$$

Since $\lambda\mu$ and $\nu\kappa$ are both half-turns around points of k, $\tau = (\lambda\mu)(\nu\kappa)$ is a translation along k, proving the result. □

19. The action of the special linear group in the trace zero model

We have seen that $\mathrm{Isom}(\mathfrak{h}) \cong \mathrm{Lor}(\mathfrak{sl}_2(\mathbb{R}))$, and we will use this to identify $\mathrm{Isom}(\mathfrak{h})$ with $\mathrm{PGL}_2(\mathbb{R}) = \mathrm{GL}_2(\mathbb{R})/\mathbb{R}^\times$. (Here $\mathbb{R}^\times$ has been identified with the subgroup of nonzero scalar matrices.) To do so, we begin by defining an action of $\mathrm{GL}_2(\mathbb{R})$ on $\mathfrak{sl}_2(\mathbb{R})$. We set

$$\sigma X = \mathrm{sign}(\sigma)\sigma X \sigma^{-1}$$

where $\sigma \in \mathrm{GL}_2(\mathbb{R})$, $X \in \mathfrak{sl}_2(\mathbb{R})$, and $\mathrm{sign}(\sigma) = \pm 1$ is the sign of the determinant of σ. We leave it to the reader to check that conjugation by σ is an orthogonal transformation of $\mathfrak{sl}_2(\mathbb{R})$, and the sign factor is there so that the action preserves sheets.

Theorem 19.1. *The action of* $\mathrm{GL}_2(\mathbb{R})$ *on* $\mathfrak{sl}_2(\mathbb{R})$ *induces an isomorphism*

$$\mathrm{PGL}_2(\mathbb{R}) \cong \mathrm{Lor}(\mathfrak{sl}_2(\mathbb{R})).$$

Proof. You have already checked that each $\sigma \in \mathrm{GL}_2(\mathbb{R})$ induces a Lorentz transformation of $\mathfrak{sl}_2(\mathbb{R})$. This means that we have a group homomorphism $f\colon \mathrm{GL}_2(\mathbb{R}) \to \mathrm{Lor}(\mathfrak{sl}_2(\mathbb{R}))$. It remains to show that f is surjective with kernel $\mathbb{R}^\times$.

For surjectivity, recall that $\mathrm{Lor}(\mathfrak{sl}_2(\mathbb{R}))$ is generated by Lorentz reflections, so it suffices to show that for every $K \in N(\mathfrak{sl}_2(\mathbb{R}))$, the reflection τ_K is in the image of f. The reader may check by calculation that for all $K, X \in \mathfrak{sl}_2(\mathbb{R})$, we have

$$KX + XK = -2\langle K, X\rangle I.$$

If $K \in N(\mathfrak{sl}_2(\mathbb{R}))$, then $K = K^{-1}$. Multiplying on the right by this factor and rearranging gives

$$-KXK^{-1} = X + 2\langle K, X\rangle K. \tag{19.2}$$

On the left, we see the action of K on X, and on the right we see $\tau_K(X)$, so f is surjective.

The determination of the kernel of f is more straightforward and we ask the reader to think through the details. Begin by assuming that σ acts trivially on $\mathfrak{sl}_2(\mathbb{R})$ and use this to show that σ commutes with every 2×2 real matrix. It follows that σ is a scalar matrix, as desired. See [Ive92, p.105] for details. □

Corollary 19.3. The action of $\mathrm{SL}_2(\mathbb{R})$ on $\mathfrak{sl}_2(\mathbb{R})$ identifies $\mathrm{PSL}_2(\mathbb{R})$ with the special Lorentz group $\mathrm{Lor}^+(\mathfrak{sl}_2(\mathbb{R}))$.

Proof. It suffices to show that the action of $\sigma \in \mathrm{GL}_2(\mathbb{R})$ is an even Lorentz transformation if and only if $\det\sigma > 0$. To this end, pick a decomposition of the action of σ on $\mathfrak{sl}_2(\mathbb{R})$ as a product $r_1, \ldots, r_s$ of reflections along vectors $K_1, \ldots, K_s \in N(\mathfrak{sl}_2(\mathbb{R}))$. It follows from (19.2) that σ and the matrix $K_1 \cdots K_s \in \mathrm{GL}_2(\mathbb{R})$ have the same action on $\mathfrak{sl}_2(\mathbb{R})$. As such, there is a scalar $\lambda \in \mathbb{R}^\times$ such that

$$\sigma = \lambda K_1 \ldots K_s.$$

Taking the determinant of both sides, we get $\det\sigma = \lambda^2(-1)^s$, so $\det\sigma > 0$ if and only if s is even. □

We can finally connect the geometric classification of the previous section to the square trace analysis we performed in Section 15. If $\sigma \in \mathrm{SL}_2(\mathbb{R})$, then its action on $\mathfrak{h}$ can be decomposed as $\alpha\beta$ where α, β are reflections in geodesics h, k, respectively. Let H and K denote normal vectors for h and k, respectively.

Theorem 19.4. *In the notation of the previous paragraph,*

$$\mathrm{tr}^2(\sigma) = 4\langle H, K\rangle^2.$$

Proof. By the proof of the preceding corollary, we can write $\sigma = \varepsilon HK$ where $\varepsilon = \pm 1$. Thus

$$\operatorname{tr}\sigma = \varepsilon \operatorname{tr}(HK) = -2\varepsilon \langle H, K \rangle .$$

Squaring both sides proves the theorem. □

This allows us to complete our dictionary between geometric and algebraic descriptions of even isometries of $\mathfrak{h}$. Geometrically, these tranformations are always the product of reflections in two geodesics, h and k. The relative position of these geodesics determines (and is determined by) the range in which the square trace $\operatorname{tr}^2 \sigma$ of the associated $\sigma \in \mathrm{PSL}_2(\mathbb{R})$, as summarized here (compare with [Ive92, Table III.4.8]):

» The geodesics h and k intersect in $\mathfrak{h}$ if and only if $0 \leq \operatorname{tr}^2(\sigma) < 4$. In this case, σ is elliptic / a rotation.
» The geodesics h and k have a common perpendicular in $\mathfrak{h}$ if and only if $\operatorname{tr}^2(\sigma) > 4$. In this case, σ is hyperbolic / a translation.
» The geodesics h and k have a common end in $\partial\mathfrak{h}$ if and only if $\operatorname{tr}^2(\sigma) = 4$. In this case, σ is parabolic / a horolation / a limit rotation.

The conjugacy class in $\mathrm{PGL}_2(\mathbb{R})$ of a general matrix $\sigma \in \mathrm{GL}_2(\mathbb{R})$ is deteremined by $\operatorname{tr}^2(\sigma)$ and $\operatorname{sign}\det\sigma$. See [Ive92, Proposition III.4.9] for details.

20. Trigonometry

Consider a hyperbolic triangle $\triangle ABC$, oriented so that its normal vectors point inwards. Let the angle at A be α, the angle at B be β, and the angle at C be γ; let the length of the side opposite A be a, opposite B be b, and opposite C be c.

Proposition 20.1 (Hyperbolic cosine and sine relations)**.** With the conventions above,

$$\cosh a = \cosh b \cosh c - \sinh b \sinh c \cos\alpha$$

and

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

Proof. Recall that

$$B = A\cosh c + V \sinh c \qquad \text{and} \qquad C = A\cosh b - U\sinh b$$

for V the unit tangent vector to AB at A and U the unit tangent to CA at A. We have $\langle U, V\rangle = \cos\alpha$, so

$$\langle B, C\rangle = \cosh b \cosh c - \sinh b \sinh c \cos\alpha.$$

But we also know that $\langle B, C\rangle = \cosh a$, and the cosine relation follows.

The sine relation follows from evaluating $\operatorname{vol}(A, B, C)$ in a symmetrical fashion. See [Ive92, II.5.1] for details. □

Proposition 20.2 (Alternative cosine relation)**.** With the conventions above,

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

Let H, K, L be the inward directed normal vectors for BC, CA, and AB, respectively. Then, by Proposition 16.3(b) and the discussion around the definition of oriented angle,

$$\langle H\wedge K, L\wedge H\rangle = \langle K, H\rangle \langle H, L\rangle - \langle H, H\rangle \langle K, L\rangle = \cos\beta\cos\gamma + \cos\alpha.$$

We also have

$$H\wedge K = C\sin\gamma \qquad \text{and} \qquad L \wedge H = B\sin\beta.$$

Combined with the definition of hyperbolic distance, we get

$$\langle H \wedge K, L \wedge H \rangle = \langle B, C \rangle \sin \gamma \sin \beta = \cosh a \sin \gamma \sin \beta.$$

Corollary 20.3. The sum of the angles in a hyperbolic triangle satisfies

$$\alpha + \beta + \gamma < \pi.$$

Proof. Without loss of generality, assume $\alpha \geq \beta, \gamma$. By the alternative cosine relation and manipulation of trig functions, $\cosh a > 1$ becomes

$$\cos(\pi - \alpha) < \cos(\beta + \gamma).$$

When $\beta + \gamma \leq \pi$, we get $\pi - \alpha > \beta + \gamma$, so $\alpha + \beta + \gamma < \pi$ as desired. If $\pi \geq \beta + \gamma$, then the display above implies $\cos(\alpha + \pi) < \cos(\beta + \gamma)$, whence $\alpha + \pi < \beta + \gamma$, contradicting $\alpha \geq \beta, \gamma$. □

Theorem 20.4. *Let $\alpha, \beta, \gamma \in (0, \pi)$ be real numbers with $\alpha + \beta + \gamma < \pi$. Then there exists a triangle $\triangle ABC$ in $\mathfrak{h}$ with $\angle A = \alpha$, $\angle B = \beta$, and $\angle C = \gamma$.*

Proof sketch. Since $\beta + \gamma < \pi - \alpha$, we know that

$$\cos(\beta + \gamma) > \cos(\pi - \alpha).$$

By angle addition formulæ, this implies

$$\sin \beta \sin \gamma < \cos \beta \cos \gamma + \cos \alpha,$$

whence there exists $a > 0$ such that

$$\sin \beta \sin \gamma \cosh a = \cos \beta \cos \gamma + \cos \alpha.$$

Now pick $B, C \in \mathfrak{h}$ with $d(B, C) = a$ and draw a geodesic ℓ through B forming angle β with BC; also draw a geodesic k through C forming angle γ with BC. Use vector calculus on normal vectors to check that ℓ and k intersect with angle α at a point A. Then $\triangle ABC$ is the desired hyperbolic triangle. See [Ive92, III.5.4] for details. □

We will now investigate the so-called *Lambert quadrilaterals*. These are quadrilaterals $\square ABCD$ with three right angles. In hyperbolic geometry, the fourth angle in such a quadrilateral is necessarily acute, as follows from the second relation below. In particular, there are no hyperbolic rectangles!

Proposition 20.5. In a Labert quadrilateral $\square ABCD$ with right angles at A, B, and D and $\angle C = \gamma$,

$$\cosh d(A, D) = \cosh d(B, c) \sin \gamma \qquad \text{and} \qquad \sinh d(A, B) \sinh d(D, A) = \cos \gamma.$$

Proof sketch. Let H, K, L, M be inward directed normal vectors for BC, CD, DA, and AB, respectively. First show that

$$\cosh d(B, C) \sin \gamma = \langle K, L \rangle$$

by proceeding as in the proof of the alternative cosine relation. Since $\langle K, L \rangle = \cosh d(A, D)$, this gives the first equality.

For the second equality, use Proposition 16.3(c) to get

$$\operatorname{vol}(K, L, M) \operatorname{vol}(M, L, H) = \det \begin{pmatrix} 0 & \langle K, L \rangle & \langle K, H \rangle \\ -1 & 0 & \langle M, H \rangle \\ 0 & -1 & 0 \end{pmatrix} = \langle K, H \rangle.$$

By the discussion at the start of §17.4 for $\langle H, K \rangle = \cos \gamma$, we get the second identity. □

Of course, $\sinh$ is positive on $(0, \infty)$ but $\cos t \leq 0$ for $\pi/2 \leq t \leq 3\pi/2$, so the fourth angle in a Lambert quadrilateral must be acute.

 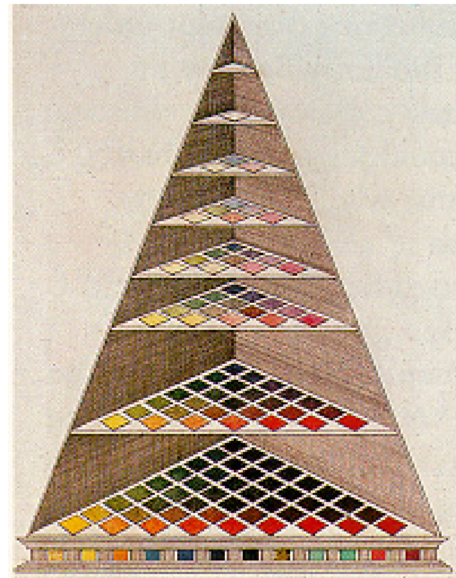

FIGURE 24. J.H. Lambert (1728–74) was a Swiss polymath who introduced hyperbolic trig functions, proved the irrationality of π, invented the hygrometer (which measures humidity), and produced one of the first three-dimensional theories of color.

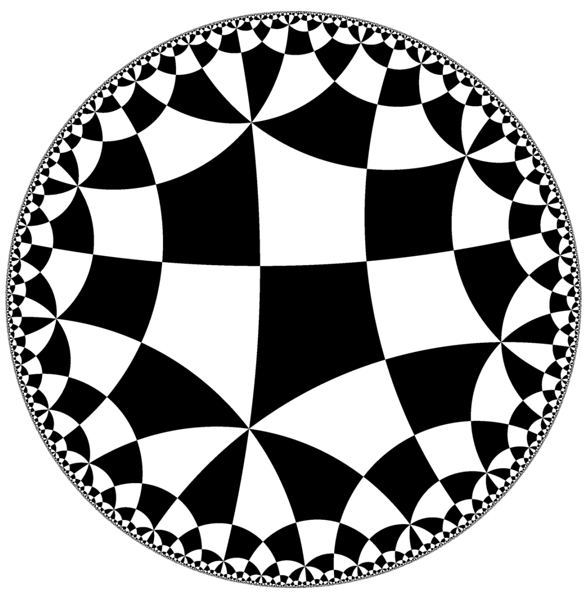

FIGURE 25. A tiling of the Poincaré disk by Lambert quadrilaterals.

21. ANGLE OF PARALLELISM

Consider a hyperbolic triangle $\triangle ABC$ in which $A \in \partial\mathfrak{h}$ is an end. Since $\triangle ABC$ is not contained in $\mathfrak{h}$, the results of the previous section do not apply. Visually, though, we might suspect that $\angle A = 0$, and the following proposition justifies this convention.

Proposition 21.1. In the above setup,

$$\cosh a \sin\beta \sin\gamma = \cos\beta\cos\gamma + 1.$$

Proof. Proceeding as we did for standard hyperbolic triangles, we still have

$$\cosh a \sin\beta \sin\gamma = \cos\beta\cos\gamma + \langle K, L\rangle .$$

Since AB and AC have a common end, $\langle K, L\rangle = \pm 1$, and the value must be $+1$ since the left-hand side above is positive. □

Note that when $\angle C = \pi/2$, the proposition implies that

$$\cosh a \sin\beta = 1,$$
$$\sinh a \tan\beta = 1,$$
$$\tanh a \sec\beta = 1$$

where the second two equalities follow from simple manipulation of the first.

22. Right-angled pentagons

We have already seen that Lambert quadrilaterals always have fourth angle acute; *i.e.*, there are no rectangles in hyperbolic geometry. Bizarrely, there are pentagons with five right angles in $\mathfrak{h}$, as we presently explore. (We prove existence in Proposition 22.3.)

Proposition 22.1. Suppose that $\pentagon P_1P_2P_3P_4P_5$ is a right-angled pentagon with $d(P_i, P_{i+1}) = a_i$ for $1 \le i \le 5$ (where we make the convention that $P_6 = P_1$). Then

$$\cosh a_5 = \sinh a_2 \sinh a_3 \qquad \text{and} \qquad \cosh a_3 = \coth a_2 \coth a_4$$

where $\coth = \cosh / \sinh$ is the hyperbolic cotangent function.

Proof. For $1 \le i \le 5$ let N_i be the inward normal vector for P_iP_{i+1}. Since all the angles are right, we know $\langle N_i, N_{i+1}\rangle = 0$ for all i. It follows from Proposition 16.3(c) that

$$\mathrm{vol}(N_1, N_2, N_3)\,\mathrm{vol}(N_2, N_3, N_4) = \langle N_1, N_4\rangle .$$

Since P_1P_2 and P_4P_5 have the common perpendicular P_5P_1, the second display in §17.4 implies that $\langle N_1, N_4\rangle = \cosh a_5$. Also by 17.4, the volume terms evaluate to $-\sinh a_2$ and $-\sinh a_3$, respectively, giving the first formula.

We defer the proof of the second identity to [Ive92, III.7]. The proof includes the more general fact that

$$\sinh a_2 \sinh a_4 \cosh a_3 = \langle N_1, N_5\rangle + \cosh a_2 \cosh a_4 \tag{22.2}$$

without the assumption that $\angle P_1 = \pi/2$. When $\angle P_1 = \pi/2$, $\langle N_1, N_5\rangle = 0$ and the identity follows. □

Proposition 22.3. Let $a_1, a_2 > 0$ be real numbers. There exists a right-angled pentagon with two consecutive sides of length a_1, a_2 if and only if $\sinh a_1 \sinh a_2 > 1$.

Proof. Necessity of the condition follows from the first formula in Proposition 22.1 since $\cosh a_5 > 1$ for $a_5 > 0$. The rest of the proof is a geometric construction. Begin by placing P_1, P_2, P_3 such that $d(P_1, P_2) = a_1$, $d(P_2, P_3) = a_2$, and $\angle P_2 = \pi/2$. Draw the geodesic k_3 through P_3 perpendicular to P_2P_3 and the geodesic k_5 through P_1 perpendicular to P_1P_2. From our study of Lambert quadrilaterals, we know that

$$\sinh a_1 \sinh a_2 = \langle N_3, N_5\rangle > 1,$$

implying that k_3 and k_5 have a common perpendicular geodesic. □

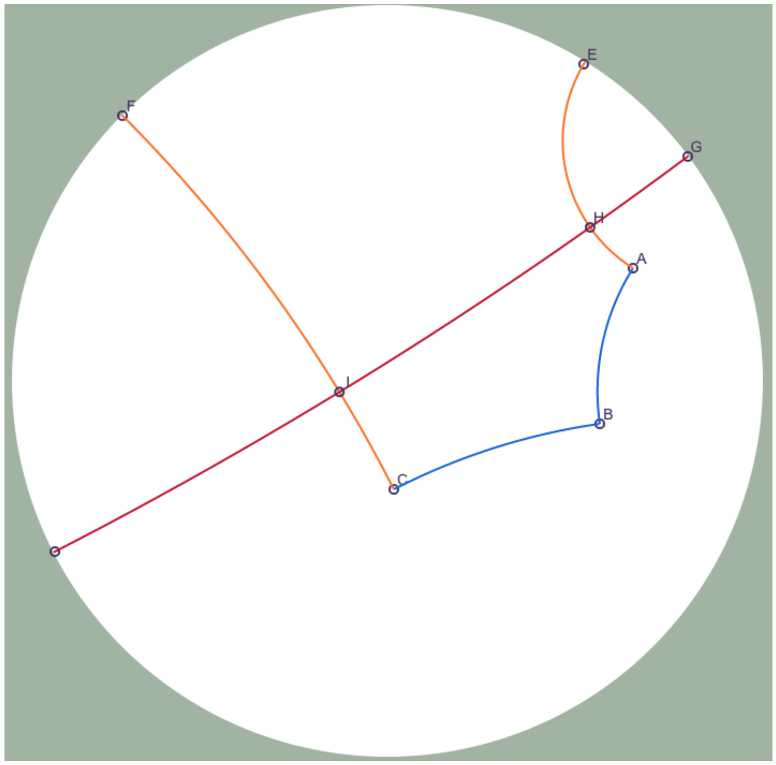

FIGURE 26. A right-angled pentagon constructed in NonEuclid via the method of Proposition 22.3. Here $A = P_1$, $B = P_2$, $C = P_3$, $k_3 = CF$, $k_5 = AE$, and IH is the common perpendicular to CF and AE.

23. RIGHT-ANGLED HEXAGONS

In future work, we will construct so-called *hyperbolic pants* by gluing together two congruent right-angled hexagons.

Proposition 23.1. Suppose that $\hexagon P_1 \cdots P_6$ is a right-angled hexagon with $d(P_i, P_{i+1}) = a_i$ for $1 \leq i \leq 6$ and $P_7 = P_1$. Then

$$\frac{\sinh a_2}{\sinh a_5} = \frac{\sinh a_6}{\sinh a_3} = \frac{\sinh a_4}{\sinh a_1}$$

and

$$\sinh a_2 \sinh a_4 \cosh a_3 = \cosh a_6 + \cosh a_2 \cosh a_4.$$

The reader is encouraged to study the proof in [Ive92, III.8]. The following theorem provides for the existence of right-angled hexagons and classifies them by the length of three alternating sides.

Theorem 23.2. *Let $a_2, a_4, a_6 > 0$ be real numbers. Then there exists a right-angled hexagon in $\mathfrak{h}$ such that three alternating sides have lengths a_2, a_4, a_6. The hexagon is unique up to isometry.*

Proof. First choose a_3 to be the unique positive real number making the second formula in Proposition 23.1 hold. Start the construction of the hexagon by placing points P_2, P_3, P_4, P_5 in the plane distances a_2, a_3, a_4 apart and with $\angle P_3 = \angle P_4 = \pi/2$. Let h_1 be the geodesic passing through P_2 at a right angle to P_2P_3, and let h_5 be the geodesic passing through P_5 at a right angle to P_5P_4. Let N_1 and N_5 denote the inward directed normal vectors for h_1 and h_5. By (22.2) and the previous proposition,

$$\langle N_1, N_5\rangle = \cosh a_6 > 1.$$

As such, h_1 and h_5 have a common perpendicular geodesic h_6. Let P_1 be the intersection of h_1 and h_6, and let P_6 be the intersection of h_5 and h_6. The above relation on normal vectors implies that $d(P_1, P_6) = a_6$. Uniqueness up to isometry is left to the reader. □

24. Hyperbolic area

We send the reader to [Ive92, III.9] to verify the following assertions.

Recall that the upper half-plane H^2 has metric

$$d(z,w) = \operatorname{arccosh}\left(1 + \frac{|z-w|^2}{2\operatorname{Im}(z)\operatorname{Im}(w)}\right).$$

Given $z = x+iy \in H^2$ with $x, y \in \mathbb{R}$ (and $y > 0$), we can build the matrix $F(z) \in \mathfrak{sl}_2(\mathbb{R})$ given by

$$F(z) = \frac{1}{y}\begin{pmatrix} x & -|z|^2 \\ 1 & -x \end{pmatrix}.$$

A calculation reveals that $F\colon H^2 \to \mathfrak{h}$ is a $\mathrm{PGL}_2(\mathbb{R})$-equivariant[11] isometry.

One may also calculate that that for $u, v \in \mathbb{R}$, $v > 0$, the derivative of F at $z = x+iy$ is the linear mapping given by

$$DF_z(u+iv) = \frac{1}{y}\begin{pmatrix} 1 & -2x \\ 0 & -1 \end{pmatrix} u - \frac{1}{y^2}\begin{pmatrix} x & -x^2+y^2 \\ 1 & -x \end{pmatrix} v.$$

As such the norm of $DF_z(u+iv)$ is

$$\det DF_z(u+iv) = -\frac{1}{y^2}(u^2+v^2).$$

Moral exericse: show that this implies that DF_z preserves angles. As such, we are justified in making the following statement.

Proposition 24.1. The transformation $F\colon H^2 \to \mathfrak{h}$ preserves oriented angles.

We can also make the calculation

$$\operatorname{vol}(F(z), DF_z(r+is), DF_z(u+iv)) = \frac{1}{y^2}(rv - su).$$

In the language of differential forms, this justifies the formula

$$F^* \operatorname{vol} = \frac{1}{y^2} dx \wedge dy,$$

whence we get the Riemannian metric

$$\frac{1}{y^2}(dx^2 + dy^2)$$

on the half-plane.[12] This means that the area of a region Δ in H^2 is

$$\operatorname{area}(\Delta) = \int_\Delta \frac{1}{y^2}\, dx\, dy.$$

We now check that the area of a hyperbolic triangle with angles α, β, γ is

$$\pi - (\alpha+\beta+\gamma). \tag{24.2}$$

[11] This means that $F(\sigma z) = \sigma \cdot F(z)$ for all $\sigma \in \mathrm{PGL}_2(\mathbb{R})$, where $\sigma \cdot F(z) = \operatorname{sign}(\sigma)\sigma F(z)\sigma^{-1}$.

[12] The reader unfamiliar with Riemannian metrics can safely ignore this claim.

In HW7, you will use this to verify that the area of a hyperbolic n-gon Δ is

$$\operatorname{area}(\Delta) = (n-2)\pi - \sum_P \angle_{\text{int}} P \tag{24.3}$$

where the sum is of interior angles at all vertices P of Δ.

Our verification of (24.2) follows [Thu97, §2.4]. First consider an ideal triangle with all vertices on ∂H^2. We may assume that the triangle has vertices ∞, $(-1,0)$, and $(1,0)$.[13] This is the region $\{x+iy \in H^2 \mid -1 \le x \le 1 \text{ and } y \ge \sqrt{1-x^2}\}$. As such, its hyperbolic area is

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

(Moral exercise: verify the above manipulations.)

Now suppose we have a "2/3-ideal triangle" with two vertices on ∂H^2 and one in H^2. Let $A(\theta)$ denote the area of such a triangle with angle $\pi - \theta$ at the finite vertex. (An isometry chase implies that this value is well-defined.) We claim that A is an additive function (proven in the following paragraph): $A(\theta_1+\theta_2) = A(\theta_1) + A(\theta_2)$ for $\theta_1, \theta_2, \theta_1+\theta_2 \in (0,\pi)$. If this is the case, then A is $\mathbb{Q}$-linear, and since it is continuous it is also $\mathbb{R}$-linear. Since $A(\pi) = \pi$ by the preceding paragraph, we learn that $A(\theta) = \theta$ and the area of a $2/3$-ideal triangle is the complement of the angle at its finite vertex.

We now verify that A is additive. (The reader is encouraged to draw a picture as they go through this argument or to view the video lecture for visual supplements.) Let $\triangle PQR$ be a $2/3$-ideal triangle with $P, Q \in \partial H^2$ and $\angle R = \pi - \theta_1$. We can construct $\triangle QRS$ with $S \in \partial H^2$ and $\angle R = \pi - \theta_2$. Let O be the intersection of the geodesics PR and QS, and let T be the end of PO which is not P. Then $\triangle TRS$ has $\angle R = \pi - (\theta_1+\theta_2)$ and area $A(\theta_1+\theta_2)$. Notice, though, that a rotation through O exhibits that $\triangle POQ \cong \triangle TOS$, whence

$$A(\theta_1+\theta_2) = A(\theta_1) + A(\theta_2).$$

Finally, we ask the reader to verify that an arbitrary hyperbolic triangle (with two or three vertices in H^2) is the difference of an ideal triangle and two or three $2/3$-ideal triangles. (Draw a picture!) The area formula (24.2) follows.

You will also show in HW7 (via an integral computation) that the area of a hyperbolic disk D of radius r is

$$\operatorname{area}(D) = 2\pi(\cosh r - 1) = 4\pi \sinh^2(r/2). \tag{24.4}$$

Note that this quantity is proportional to e^r as $r \to \infty$, in stark contrast with the area of Euclidean disks (proportional to r^2).

25. Fuchsian groups

A discrete subgroup of $\mathrm{PSL}_2(\mathbb{R}) \cong \mathrm{Isom}^+(\mathbb{H}^2)$ is called a *Fuchsian group*. Presently, we study slightly more general objects, namely discrete subgroups of $\mathrm{PGL}_2(\mathbb{R})$.

25.1. **Discrete subgroups of locally compact groups.**

In order to initiate the study of Fuchsian groups, we start with a few words on locally compact topological groups and their discrete subgroups. Students who have not studied point-set topology previously can focus on the theorem statements below. The important topological notions are the following:

[13] Moral exercises: (1) Find isometries taking an arbitrary ideal triangle to this one. (2) Prove that isometries preserve area.

- *topological space*: a set equipped with a class of open subsets satisfying some axioms (empty and total subsets are open, opens are closed under arbitrary unions, and opens are closed under finite intersections); closed sets are (by definition) the complements of open sets;
- *continuous function*: a function $f\colon X \to Y$ between topological spaces such that the preimage of every open set in Y is open in X;
- *[adjective] neighborhood of a point*: a subset of the ambient space of [adjective] type containing the point;
- *Hausdorff space*: a topological space for which every two points have disjoint open neighborhoods;
- *locally compact space*: a Hausdorff space in which every point has a compact neighborhood;
- *isolated point x of $S \subseteq X$*: the point x has a neighborhood V in X such that $V \cap S = \{x\}$;
- *discrete subset $S \subseteq X$*: every point of S is isolated;
- *accumulation point*: a point $w \in X$ is an accumulation point for $S \subseteq X$ if every open neighborhood of w contains infinitely many points of S.

We state without proof the following theorem on closed discrete subsets. (Note that $\{1/n \mid n \in \mathbb{Z}^+\}$ is a discrete subset of $\mathbb{R}$ which is not closed.)

Theorem 25.1. *The following statements about a subset S of a topological space X are equivalent:*

(1) S is closed and discrete,

(2) S has no accumulation points.

If, additionally, X is locally compact, then it is also equivalent that

(3) S intersects every compact subset of X in a finite set.

Henceforth, G will denote a locally compact topological group with identity element e. This means that G is a locally compact space and that the product $G \times G \to G$ and inverse $G \to G$ functions are continuous. A *discrete subgroup* $\Gamma \leq G$ is a subgroup of G which is a discrete subset of G.

Proposition 25.2. Let $\Gamma \leq G$ be a subgroup of G. If there exists an open neighborhood V of $e \in G$ such that $V \cap \Gamma = \{e\}$, then Γ is a discrete subgroup of G.

Proof. Let V be such a neighborhood of e. Given $\sigma \in G$, we aim to construct an open neighborhood U of σ in G such that $U \cap \Gamma = \{\sigma\}$ or $\varnothing$. We either have

$$\sigma(V^{-1}) \cap \Gamma = \varnothing \qquad \text{or} \qquad \sigma(V^{-1}) \cap \Gamma \neq \varnothing$$

where $V^{-1} = \{v^{-1} \mid v \in V\}$. In the first case, $\sigma(V^{-1})$ is a neighborhood of σ disjoint from Γ. In the second case, we may take $\gamma \in \sigma(V^{-1}) \cap \Gamma$. Then $\sigma \in \gamma V$, $\gamma \in \Gamma$, and

$$\gamma^{-1}(\gamma V \cap \Gamma) = V \cap \gamma^{-1}\Gamma = V \cap \Gamma = \{e\}.$$

We conclude that γV is an open neighborhood of σ meeting Γ in the unique point γ, as desired. □

We say that a group action of G on a topological space X is *continuous* if the associated function $G \times X \to X$ is continuous.

Corollary 25.3. Suppose G acts continuously on X and let Γ be a subgroup of G. If there exists a point $x \in X$ which is isolated in the orbit Γx and such that the stabilizer $\Gamma_x = \{\gamma \in \Gamma \mid \gamma x = x\}$ is discrete in G, then Γ is a discrete subgroup of G.

Proof. Let U be an open neigborhood of x in X such that $U \cap \Gamma x = \{x\}$. Then the inverse image of U along the map $\tau_x\colon G \to X$, $g \mapsto gx$ satisfies

$$\tau_x^{-1}(U) \cap \Gamma = \Gamma_x.$$

Pick an open neighborhood V of e in G such that $V \cap \Gamma_x = \{e\}$. Then $V \cap \tau_x^{-1}(U)$ isolates e in Γ, and Proposition 25.2 applies. □

25.2. **Discrete subgroups of the projective general linear group.** In order to study Fuchsian groups, we will need to specify a topology on $\mathrm{PGL}_2(\mathbb{R})$. To do so, first note that $M_{2\times 2}(\mathbb{R}) \cong \mathbb{R}^4$ has the standard Euclidean topology, and $\mathrm{GL}_2(\mathbb{R})$ — the complement of the closed set defined by $ad - bc = 0$ — is an open subset of $M_{2\times 2}(\mathbb{R})$. A set $U \subseteq \mathrm{GL}_2(\mathbb{R})$ is open if it is open in $M_{2\times 2}(\mathbb{R})$. (Equivalently, open sets in $\mathrm{GL}_2(\mathbb{R})$ are of the form $V \cap \mathrm{GL}_2(\mathbb{R})$ where $V \subseteq M_{2\times 2}(\mathbb{R})$ is open.) The quotient group $\mathrm{PGL}_2(\mathbb{R}) = \mathrm{GL}_2(\mathbb{R})/\mathbb{R}^\times$ carries the natural quotient topology. Its open sets correspond to open subsets of $\mathrm{GL}_2(\mathbb{R})$ that are stable under the action of $\mathbb{R}^\times$.

We will use the action of $\mathrm{PGL}_2(\mathbb{R})$ on H^2 to study its discrete subgroups, but first it will be convenient to produce some compact neighborhood of $e \in \mathrm{PGL}_2(\mathbb{R})$.

Theorem 25.4. *Given $z \in H^2$ and $\varepsilon > 0$, let*

$$\mathscr{R}(z;\varepsilon) = \{\sigma \in \mathrm{PGL}_2(\mathbb{R}) \mid d(z, \sigma z) \le \varepsilon\}.$$

Then $\mathscr{R}(z;\varepsilon)$ is a compact neighborhood of e.

Proof. Define the norm $\|\sigma\|$ of a matrix $\sigma \in \mathrm{GL}_2(\mathbb{R})$ via

$$\left\| \begin{pmatrix} a & b \\ c & d \end{pmatrix} \right\| = \sqrt{\frac{a^2+b^2+c^2+d^2}{|\det \sigma|}}.$$

Since $\|\lambda\sigma\| = \|\sigma\|$ for $\lambda \in \mathbb{R}^\times$, the norm descends to a real-valued function on $\mathrm{PGL}_2(\mathbb{R})$, and it is not hard to check that it is continuous. As such, the set

$$\{\sigma \in \mathrm{PGL}_2(\mathbb{R}) \mid \|\sigma\| \le r\} \tag{25.5}$$

is a neighborhood of e in $\mathrm{PGL}_2(\mathbb{R})$ for all $r > \sqrt{2}$. This neighborhood is compact because it is the image of a closed and bounded set in $\mathrm{GL}_2(\mathbb{R}) \subseteq M_{2\times 2}(\mathbb{R}) \cong \mathbb{R}^4$.

We now show that $\mathscr{R}(i;\varepsilon)$ is a compact neighborhood of e, where $i = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \in \mathfrak{h}$. A calculation (see [Ive92, p.126]) reveals that $2\cosh d(i, \sigma i) = \|\sigma\|$ for $\sigma \in \mathrm{PGL}_2(\mathbb{R})$. Since sets of the form (25.5) are compact neighborhoods of e in $\mathrm{PGL}_2(\mathbb{R})$, we learn that the set $\mathscr{R}(i;\varepsilon)$ is a compact neighborhood of e as well.

Given an arbitrary point $z \in \mathfrak{h}$, transitivity of $\mathrm{PGL}_2(\mathbb{R})$ acting on $\mathfrak{h}$ implies that there exists $\tau \in \mathrm{PGL}_2(\mathbb{R})$ such that $\tau z = i$, so

$$\begin{aligned} \mathscr{R}(z;\varepsilon) &= \{\sigma \mid d(\tau^{-1} i, \sigma\tau^{-1} i) \le \varepsilon\} \\ &= \{\sigma \mid d(i, \tau\sigma\tau^{-1} i \le \varepsilon\} \\ &= \tau^{-1}\mathscr{R}(i;\varepsilon)\tau. \end{aligned}$$

We conclude that $\mathscr{R}(z;\varepsilon)$ is a compact neighborhood of e, as desired. □

Corollary 25.6. Let $\Gamma \le \mathrm{PGL}_2(\mathbb{R})$. If Γ is discrete, then for all $z \in \mathfrak{h}$ and $\varepsilon > 0$ the set

$$\{\sigma \in \Gamma \mid d(z, \sigma(z)) \le \varepsilon\}$$

is finite. Conversely, if for one $\varepsilon > 0$ and one point $z \in \mathfrak{h}$ the above set is finite, then Γ is a discrete subgroup of $\mathrm{PGL}_2(\mathbb{R})$.

Proof. The set in question is $\mathscr{R}(z;\varepsilon) \cap \Gamma$. The result follows from Theorem 25.4 and Proposition 25.2. □

Example 25.7. Let $\mathrm{GL}_2(\mathbb{Z})$ denote the group of 2×2 integer matrices with determinant ± 1. This is a discrete subgroup of $\mathrm{GL}_2(\mathbb{R})$ since any of the subsets

$$\{\sigma \in \mathrm{GL}_2(\mathbb{Z}) \mid \|\sigma\| \le n\}$$

is finite (and this is the intersection of $\mathrm{GL}_2(\mathbb{Z})$ with a compact neighborhood of e for $n > \sqrt{2}$). It is a general fact that any discrete subgroup $\Gamma \leq \mathrm{GL}_2(\mathbb{R})$ consisting of matrices with determinant ± 1 maps onto a discrete subgroup of $\mathrm{PGL}_2(\mathbb{R})$, so $\mathrm{PGL}_2(\mathbb{Z}) = \mathrm{GL}_2(\mathbb{Z})/\{\pm I\}$ is discrete subgroup of $\mathrm{PGL}_2(\mathbb{R})$.

Proposition 25.8. A discrete subgroup $\Gamma \leq \mathrm{PGL}_2(\mathbb{R})$ acts on $\mathfrak{h}$ with discrete orbits.

Proof. Let Γz be an orbit. By Corollary 25.6, a disk D with center z will contain finitely many points from Γz. Thus there exists $\varepsilon > 0$ such that $D(z;\varepsilon) \cap \Gamma z = \{z\}$. We now show that in fact $d(u,v) \geq \varepsilon$ for all $u \neq v \in \Gamma z$, implying that Γz is discrete. To see this, choose $\sigma \in \Gamma$ with $\sigma z = u$. Then $\sigma v \notin D(z;\varepsilon)$ so $d(u,v) = d(\sigma u, \sigma v) \geq \varepsilon$. □

Remark 25.9. You will show in HW7 that there is a natural action of $\mathbb{R}$ on the circle $\mathbb{R}/2\pi\mathbb{Z}$ for which $\mathbb{Z}$-orbits are not necessarily discrete, so the above proposition is special to the action of discrete subgroups of $\mathrm{PGL}_2(\mathbb{R})$ on $\mathfrak{h}$.

Theorem 25.10. *Let $\Gamma \leq \mathrm{PGL}_2(\mathbb{R})$ be discrete. For compact subsets $K, L \subseteq \mathfrak{h}$, the sets $\sigma(K)$ and L are disjoint for all but finitely many $\sigma \in \Gamma$.*

Proof. Given a compact set $K \subseteq \mathfrak{h}$, a point $z \in \mathfrak{h}$, and a number $r > 0$, it suffices to show that $D(z;r)$ and $\sigma(K)$ are disjoint for all but finitely many $\sigma \in \Gamma$. (Moral exercise: why does this suffice?) Choose $s > 0$ with $K \subseteq D(z; r+s)$ and conclude from Corollary 25.6 that $S = \{\gamma \in \Gamma \mid \gamma(z) \in D(z; r+s)\}$ is finite. Observe that $\gamma D(z;r) = D(\gamma(z); r)$ and conclude that $K \cap \gamma D(z;r) = \varnothing$ for all $\gamma \in \Gamma \smallsetminus S$. □

Remark 25.11. The property described in the theorem has a name: the action of $\mathrm{PGL}_2(\mathbb{R})$ on $\mathfrak{h}$ is *properly discontinuous*. This puts us in the unfortunate terminological bind of working with a continuous properly discontinuous action, but such is life. Some authors use the term *properly discrete* instead, but this is far less common. When a locally compact group G acts continuously on a space X in a properly discontinuous fashion, the quotient map $X \to X/G$ is a covering map. This leads other authors (for example, Hatcher) to call such actions *covering actions*.

Corollary 25.12. Let Γ be a discrete subgroup of $\mathrm{PGL}_2(\mathbb{R})$ and K a compact subset of $\mathfrak{h}$. Only finitely many elements of Γ have fixed points in K.

Proof. For all but finitely many $\sigma \in \Gamma$, the sets K and $\sigma(K)$ are disjoint. □

An *elliptic fixed point* for Γ on $\mathfrak{h}$ is a point $z \in \mathfrak{h}$ which is fixed by some proper elliptic element $\sigma \in \Gamma$. (Recall that elliptic means $\sigma \in \mathrm{PSL}_2(\mathbb{R})$ with $\operatorname{tr}^2 \sigma < 4$; these correspond to hyperbolic rotations.)

Corollary 25.13. Let $\Gamma \leq \mathrm{PGL}_2(\mathbb{R})$ be discrete. The set of elliptic fixed points for Γ on $\mathfrak{h}$ is a discrete subset of $\mathfrak{h}$.

Proof. This follows immediately from the previous corollary and Theorem 25.1 as soon as the reader has checked that every Fuchsian group is closed. □

In fact, a converse to Corollary 25.13 is true as well. Call $\Gamma \leq \mathrm{PGL}_2(\mathbb{R})$ *elementary* if it fixes a point of $\mathfrak{h}$, fixes a point of $\partial\mathfrak{h}$, or stabilizes a geodesic in $\mathfrak{h}$. It is a fact (proved in [Ive92, §IV.2]) that subgroups of $\mathrm{PGL}_2(\mathbb{R})$ are elementary if and only if they are solvable.

Theorem 25.14 (Jakob Nielsen's theorem)**.** *A non-elementary subgroup Γ of $\mathrm{PGL}_2(\mathbb{R})$ is discrete if and only if the set of elliptic fixed points in $\mathfrak{h}$ is discrete.*

The proof of this theorem is difficult and occupies §§IV.2-4 of [Ive92]. Due to time constraints, we will not present it here.

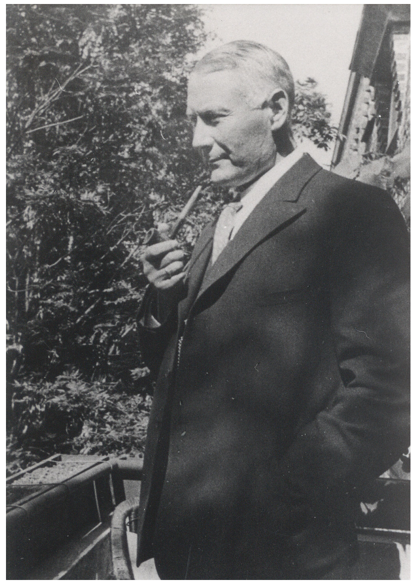

FIGURE 27. Jakob Nielsen (1890–1959) was a Danish mathematician who pioneered geometric group theory. From 1952 to 1958, he was on the executive board of UNESCO.

26. CUSPS AND HOROCYCLIC COMPACTIFICATION

Note to the reader: Because of time constraints, these notes are going to switch to a more descriptive style. See the corresponding portions of [Ive92] for full details and proofs. If you would like to see some elaboration on any portion of these notes or the textbook, please say so on the Slack channel.

We now study how a discrete group $\Gamma \leq \mathrm{PGL}_2(\mathbb{R})$ acts on the boundary of the hyperbolic plane. A point $S \in \partial\mathfrak{h}$ is a *cusp* for Γ if it is fixed by a proper horolation (or parabolic transformation) $\beta \in \Gamma$. (By proper, we mean not the identity element.) The stabilizer Γ_S of a cusp consists of horolations with center S and reflections in geodesics through S. A *horodisk* is a region in $\mathfrak{h}$ bounded by a horocycle. (Recall that horocycles are the orbits of the group of horolations with center S. In the upper half-plane model, they are generalized circles in $H^2 \cup \partial H^2$ parallel to ∂H^2 at S. When $S = \infty$, this is just a horizontal line in $\mathfrak{h}$ bounding the horodisk consisting of points with imaginary part greater than some constant.)

Proposition 26.1. Given a cusp $S \in \partial\mathfrak{h}$ for a discrete group $\Gamma \leq \mathrm{PGL}_2(\mathbb{R})$, there exists a horodisk D such that $\gamma(D) \cap D = \varnothing$ for $\gamma \in \Gamma \smallsetminus \Gamma_S$.

Similarly, we can choose a horodisk at a cusp so that its Γ-orbits avoid any compact subset of $\mathfrak{h}$:

Proposition 26.2. Given a cusp $S \in \partial\mathfrak{h}$ for a discrete group $\Gamma \leq \mathrm{PGL}_2(\mathbb{R})$ and a compact set $K \subseteq \mathfrak{h}$, there exists a horodisk D with center S such that $\gamma(D) \cap K = \varnothing$ for all $\gamma \in \Gamma$.

Corollary 26.3. Given cusps $S, T \in \partial\mathfrak{h}$ for a discrete group $\Gamma \leq \mathrm{PGL}_2(\mathbb{R})$ which are in different Γ-orbits, for any horodisk E with center T there exists a horodisk D with center S such that $\gamma(D) \cap E = \varnothing$ for all $\gamma \in \Gamma$.

We will be very interested in the quotient spaces $\mathfrak{h}/\Gamma$ for $\Gamma \leq \mathrm{PGL}_2(\mathbb{R})$ discrete. By definition, $\mathfrak{h}/\Gamma$ is the set of Γ-orbits of points in $\mathfrak{h}$ with the topology defined by the condition that $U \subseteq \mathfrak{h}/\Gamma$ is open if and only if $q^{-1}U \subseteq \mathfrak{h}$ is open for $q \colon \mathfrak{h} \to \mathfrak{h}/\Gamma$ the quotient map. Observe that $q^{-1}(q(U)) = \bigcup_{\gamma \in \Gamma} \gamma(U)$, so q takes open sets to open sets.[14]

[14] This makes q an *open map*. Beware that not all cotinuous maps are open!

Corollary 26.4. If $\Gamma \leq \mathrm{PGL}_2(\mathbb{R})$ is discrete and $\mathfrak{h}/\Gamma$ is compact, then the group Γ contains no proper horolations.

Proof. Cover $\mathfrak{h}$ with open disks. The image of these disks is an open cover of $\mathfrak{h}/\Gamma$. By compactness, we can take a finite subcover. Lifting back to $\mathfrak{h}$, we see that there is a finite set of disks in $\mathfrak{h}$ meeting all Γ-orbits. As such, we can find a compact set $K \subseteq \mathfrak{h}$ meeting all Γ-orbits (the union of the disks). Corollary 26.3 implies that Γ contains no proper horolations. □

For the rest of this section, fix a discrete group $\Gamma \leq \mathrm{PGL}_2(\mathbb{R})$ and let $Y = Y(\Gamma)$ be the subset of $\overline{\mathfrak{h}} = \mathfrak{h} \cup \partial\mathfrak{h}$ consisting of $\mathfrak{h}$ and the cusps for Γ. We give Y the following topology: $W \subseteq Y$ is open if for all $S \in W$ there exists a disk or horodisk with center S entirely contained in W. This is called the *horocyclic topology*.

Since Γ acts continuously on Y, the orbit space $X = Y/\Gamma$ is a topological space which is in fact a topological surface (two-dimensional manifold) with boundary:

Theorem 26.5. *The space $X = Y/\Gamma$ is Hausdorff and every point has a neighborhood homeomorphic to an open neighborhood of 0 in $\mathbb{R}^2$ or an open neighborhood of 0 in $\mathbb{R} \times [0, \infty)$.*

Proof. That X is Hausdorff follows from Proposition 26.2 and Corollary 26.3. We separately construct neighborhoods of cusp and regular points. To construct a neighborhood of a cusp point, it suffices to consider the point $\infty \in H^2$ and $\Gamma = \Gamma_\infty$ (by Proposition 26.1). If $\Gamma = \Gamma^+$ is generated by $z \mapsto z + k$ for some $k > 0$, then Y/Γ is homeomorphic to the open unit disk D in the complex plane via

$$\begin{aligned} Y/\Gamma &\longrightarrow D \\ z &\longmapsto \exp(2\pi i z/k). \end{aligned}$$

In the general case, we can take Γ to be generated by $z \mapsto -z\overline{z}$ and $z \mapsto z + k$, whence X is the orbit space for the action of complex conjugation on the complex unit disk in $\mathbb{C}$.

For a regular point, it suffices to consider the origin 0 in the Poincaré disk and assume $\Gamma = \Gamma_0$. If $\Gamma = \Gamma^+$ is generated by $z \mapsto \theta z$ for θ a primitive n-th root of unity, then the fact that $z \mapsto z^n$ is open implies that $H^2/\Gamma \cong D$. In the case that Γ/Γ^+ is nontrivial, a nonidentity element corresponds to a reflection of D in a line through 0, so we again get a half-space neighborhood. □

Corollary 26.6. The space $X = Y/\Gamma$ is compact if and only if there exists a compact subset $K \subseteq \mathfrak{h}$ and a finite number of horodisks $D_1, \ldots, D_r$ with centers $S_1, \ldots, S_r$ cusps for Γ such that any Γ-orbit in $\mathfrak{h}$ meets $K \cup D_1 \cup \cdots D_r$.

When Γ is a Fuchsian group (a discrete subgroup of $\mathrm{PSL}_2(\mathbb{R})$), the proof of Theorem 26.5 implies that $X = Y/\Gamma$ is a *Riemann surface*, that is, a one-dimensional complex-analytic manifold. When X is compact, it is called the *horocyclic compactification* of $\mathfrak{h}/\Gamma$.

27. The modular group and its fundamental domain

Consider the action of $(\mathbb{R}, +)$ on itself by translation. Let $\Gamma \leq \mathbb{R}$ be the discrete group $\mathbb{Z}$. To understand $\mathbb{R}/\mathbb{Z}$, it is convenient to select a "nice" set of orbit representatives. Let $U = (0, 1)$. Then U and $n + U$ are disjoint for all $n \in \mathbb{Z} \setminus \{0\}$, and each $\mathbb{Z}$-orbit meets $\overline{U} = [0, 1]$. Thus $\mathbb{R}/\mathbb{Z}$ has the same topology as $\overline{U}/(0 \sim 1)$, which is clearly a model for the circle. The set U is a fundamental domain for the action of $\mathbb{Z}$ on $\mathbb{R}$.

Definition 27.1. A *fundamental domain* U for a group $\Gamma \leq \mathrm{PGL}_2(\mathbb{R})$ is an open subset of $\mathbb{H}^2$ such that U and γU are disjoint for $\gamma \in \Gamma \setminus \{e\}$ and such that each Γ-orbit meets the closure $\overline{U}$ of U.

We now construct a fundamental domain for the *modular group* $\mathrm{PSL}_2(\mathbb{Z})$.

Proposition 27.2. The three geodesics in the upper half-plane H^2 defined by

$$\mathrm{Re}(z) = \frac{1}{2}, \qquad \mathrm{Re}(z) = -\frac{1}{2}, \qquad \text{and} \qquad |z| = 1$$

bound an open set M which is a fundamental domain for the group $\mathrm{PSL}_2(\mathbb{Z})$.

Proof. Let $\sigma, \tau \in \mathrm{Isom}^+(H^2)$ be the transformations $z \mapsto -z^{-1}$ and $z \mapsto z+1$, respectively. The corresponding matrices are $S = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ and $T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$. Let $\omega = \tau^{-1} \colon z \mapsto z - 1$ with matrix $T^{-1} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}$. Fix $w = yi$ for some $y > 1$ and note that the geodesics of the proposition are the perpendicular bisectors for $[w, \tau(w)]$, $[w, \omega(w)]$, and $[w, \sigma(w)]$, respecitvely. As such, we have

$$\overline{M} = \{z \in H^2 \mid d(w,z) \leq d(w, \gamma(z)) \text{ for } \gamma = \sigma, \tau, \omega\}$$

We currently aim to find a point of $\overline{M}$ in the orbit of a given point $z \in H^2$. Choose $\nu \in \mathrm{PSL}_2(\mathbb{Z})$ such that $d(w, \nu(z)) \leq d(w, \mu(z))$ for all $\mu \in \mathrm{PSL}_2(\mathbb{Z})$. (Important moral exercise: How do you guarantee the existence of such ν?) In particular, for any of $\gamma = \sigma$, τ, or ω, we get

$$d(w, \nu(z)) \leq d(w, \gamma(z))$$

so $\nu(z) \in \overline{M}$.

It remains to show that any orbit $\mathscr{O}$ of $\mathrm{PSL}_2(\mathbb{Z})$ meets M in at most one point. We claim that the function $\mathscr{O} \to \mathbb{R}$, $z \mapsto \mathrm{Im}(z)$ is bounded and attains its upper bound when $z \in \overline{M}$. Indeed, the reader may check that for $z \in \overline{M}$ and $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z})$,

$$\mathrm{Im}\left(\frac{az+b}{cz+d}\right) = \frac{\mathrm{Im}(z)}{|cz+d|^2}$$

and $|cz+d| > 1$. Thus $\mathrm{Im}(\mu(z)) \leq \mathrm{Im}(z)$ for all $z \in \overline{M}$ and $\mu \in \mathrm{PSL}_2(\mathbb{Z})$. Furthermore, the inequality is strict when $c \neq 0$, and when $c = 0$ the transformation is of the form $z \mapsto z + b$, $b \in \mathbb{Z}$.

Suppose now that $z, w \in \mathscr{O} \cap M$. By the above argument (and the fact that z and w are both $\mathrm{PSL}_2(\mathbb{Z})$-translates of each other), we learn that $\mathrm{Im}(z) = \mathrm{Im}(w)$. But then $w = z + b$ for some $b \in \mathbb{Z}$ (and still $z, w \in M$) so $b = 0$ and $z = w$. This verifies that $|\mathscr{O} \cap M| \leq 1$. □

It is in fact the case that $M \cup \{z \in \partial M \mid \mathrm{Re}(z) \geq 0\}$ is a set of orbit representatives of $\mathrm{PSL}_2(\mathbb{Z})$ acting on H^2. We can think of $H^2/\,\mathrm{PSL}_2(\mathbb{Z})$ as $\overline{M}$ with the left vertical glued to the right vertical (by τ) and the left half of the arc glued to the right half (via σ).[15]

Later, we will see that $\mathrm{PSL}_2(\mathbb{Z})$ is generated by σ and τ. The transformation $\rho = \tau\sigma$ is a rotation by $2\pi/3$, so we have the relations

$$\sigma^2 = e = \rho^3.$$

It is a consequence of Poincaré's theorem (one of our end goals for the course) that this set of relations is complete, that is

$$\mathrm{PSL}_2(\mathbb{Z}) \cong \left\langle S, T \mid S^2, (TS)^3 \right\rangle$$

in the language of group presentations.

Now consider the *extended modular group* $\mathrm{PGL}_2(\mathbb{Z})$. The reflection β in the imaginary axis divides M into two halves, and it is in fact the case that the hyperbolic triangle Δ with vertices

[15] To make this completely rigorous, we will need to observe that M is a *locally finite* fundamental domain. This is covered in the next section.

$i, \zeta = e^{2\pi i/3}, \infty$ is a fundamental domain for $\mathrm{PGL}_2(\mathbb{Z})$. Let α denote reflection in the left vertical side $[\zeta, \infty]$ of Δ and let γ denote reflection in $[\zeta, i]$. One can then geometrically deduce the identities

$$\alpha^2 = \beta^2 = \gamma^2 = e, \quad \alpha\gamma\alpha = \gamma\alpha\gamma, \quad \text{and} \quad \beta\gamma = \gamma\beta$$

in $\mathrm{PGL}_2(\mathbb{Z})$! (Again, Poincaré's theorem will tell us that these relations are complete.)

Lemma 27.3. Let Γ denote a discrete group of isometries of H^2 and suppose $\Pi \leq \Gamma$ is a subgroup of finite index. If D is a fundamental domain for Γ and $S \subseteq \Gamma$ is a full set of representatives for $\Pi\backslash\Gamma$, then the interior U of

$$F = \bigcup_{\alpha \in S} \alpha\overline{D}$$

is a fundamental domain for Π.

Proof. See [Ive92, Lemma V.1.4]. I'll draw a picture in lecture. □

We will use this lemma to describe a fundamental domain for the *level* 2 *modular group* $\Gamma(2)$. We begin with the group $G(2)$ which is the kernel of the surjective homomorphism $\mathrm{PGL}_2(\mathbb{Z}) \to \mathrm{PGL}_2(\mathbb{F}_2)$ which reduces the entries in an integer matrix mod 2. Since $1 = -1$ in $\mathbb{F}_2 = \mathbb{Z}/2\mathbb{Z}$, we have $\mathrm{PGL}_2(\mathbb{F}_2) = \mathrm{GL}_2(\mathbb{F}_2)$. The action of $\mathrm{GL}_2(\mathbb{F}_2)$ on the three lines in $\mathbb{F}_2^2$ reveals that it is isomorphic to the permutation group Σ_3. We can lift this group back to $\mathrm{PGL}_2(\mathbb{Z})$ as the dihedral group D_6 generated by α and γ, which have matrices

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

respectively. By the lemma, the union of the six translates of Δ by D_6 is a fundamental domain for $G(2)$. As such, the ideal hyperbolic triangle with vertices $0, -1, \infty$ is a fundamental domain for $G(2)$.

The group $\Gamma(2)$ is $G(2)^+$, the even isometries in $G(2)$. Applying the lemma to the reflection β, we see that the ideal quadrilateral with vertices $1, 0, -1, \infty$ is a fundamental domain for $\Gamma(2)$.

28. Locally finite and convex fundamental domains

Throughout this section, let Γ be a discrete subgroup of $\mathrm{PGL}_2(\mathbb{R})$. In order for a fundamental domain to provide good control over the geometry of $\mathbb{H}^2/\Gamma$, it must additionally be locally finite and convex. We will state the results here, but not pursue the proofs. See [Ive92, §§V.2–3] for details.

A fundamental domain $P \subseteq \mathbb{H}^2$ for Γ is *locally finite* if every $z \in \mathbb{H}^2$ has a neighborhood which meets $\sigma(P)$ for most finitely many $\sigma \in \Gamma$. Constructing a non-locally finite fundamental domains require a certain perversity of thought and we will not pursue an example here. Beware, though, that they exist!

When P is locally finite, we have the following two crucial results.

Proposition 28.1. If $P \subseteq \mathbb{H}^2$ is a locally finite fundamental domain for Γ, then the canonical projection map $p\colon \overline{P}/\Gamma \to \mathbb{H}^2/\Gamma$ is a homeomorphism.

Corollary 28.2. If $P \subseteq \mathbb{H}^2$ is a locally finite fundamental domain for Γ, then $\mathbb{H}^2/\Gamma$ is compact if and only if P is bounded.

A region $R \subseteq \mathbb{H}^2$ is *convex* if for all $A, B \in R$, the geodesic arc $[A, B]$ is contained in R. It turns out that a convex locally finite fundamental domain P is a hyperbolic polygon: its sides are of the form $\overline{P} \cap \sigma\overline{P}$, $\sigma \in \Gamma \smallsetminus \{e\}$; a vertex of P is of the form $\overline{P} \cap \sigma\overline{P} \cap \tau\overline{P}$, $\sigma \neq \tau \in \Gamma \smallsetminus \{e\}$. The boundary of P is the union of these sides, and the vertices are necessarily endpoints of exactly two sides.

For the remainder of this section, fix P a convex locally finite fundamental domain for Γ. A side of P has a presentation as $s = \overline{P} \cap \sigma_s \overline{P}$ for a unique $\sigma_s \in \Gamma \smallsetminus \{e\}$. We define the $*$-operator on sides, called the *side pairing*, by

$$*s = \overline{P} \cap \sigma_s^{-1} \overline{P}.$$

Then $\sigma_s(*s) = s$, and $s \mapsto *s$ is an involution on the set of sides of P.

We will call an oriented side of P an *edge*. We can extend $*$ in a natural way to the set $\mathscr{E}$ of edges of P. Let $\mathscr{E}_+$ be the set of edges with initial vertex in $\mathbb{H}^2$ (as opposed to $\partial\mathbb{H}^2$). Given an edge s with initial vertex A, let $\downarrow s$ denote the second edge with initial vertex A. We define a new operator

$$\begin{aligned} \Psi \colon \mathscr{E}_+ &\longrightarrow \mathscr{E}_+ \\ s &\longmapsto \downarrow *s. \end{aligned}$$

Since Ψ is the composition of two involutions, it is a bijection.

Example 28.3. Let s be the $\mathrm{Re}(z) = -1/2$ side of the modular fundamental domain M for $\mathrm{PSL}_2(\mathbb{Z})$, oriented so that $\zeta = e^{2\pi i/3}$ is the initial vertex. Then $\sigma_s = \omega = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}$ with $\sigma_s^{-1} = \tau = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$. Thus $*s = \overline{M} \cap \tau\overline{M} = t$ where t is the $\mathrm{Re}(z) = 1/2$ side of M, oriented so that $-\zeta^{-1}$ is its initial vertex. Thus $\Psi s = \downarrow *s$ is the oriented geodesic arc $u = [-\zeta^{-1}, i]$.

Lemma 28.4. The operator Ψ has finite orbits.

Proof. For $s \in \mathscr{E}_+$ with initial vertex A, the sequence of edges $s, \Psi s, \Psi^2 s, \Psi^3 s$ has initial vertices in $P \cap \Gamma A$, a finite set. □

Example 28.5. Continuing Example 28.3, one see that $*u = v = [\zeta, i]$ so $\Psi u = \downarrow v = s$. Thus the Ψ-orbit of s is just $\{s, u\}$. Similarly, one has $\Psi v = t$ and $\Psi t = v$.

Lemma 28.6. Let s_0 be an edge of P with initial vertex A and let $s_i = \Psi^i s_0$. Set $\sigma_i = \sigma_{s_i}$. Then there exists $n \geq 1$ with $\sigma_0\sigma_1 \cdots \sigma_n = e$ and such that

$$\bigcup_{0 \leq i \leq n} \sigma_0\sigma_1 \cdots \sigma_i \overline{P}$$

is a neighborhood of A and

$$\bigcup_{0 \leq i \leq n} \sigma_0\sigma_1 \cdots \sigma_i \overline{P}$$

is a disjoint union.

Before discussing the proof, let's see how this lemma plays out for $P = M$.

Example 28.7. If $s_0 = s$, then $s_1 = u$, $\sigma_0 = \omega$, and $\sigma_1 = \sigma = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$. The reader may check that $\omega\sigma$ is a rotation by $2\pi/3$ about ζ, and $\sigma_0\sigma_1 \cdots \sigma_5 = (\omega\sigma)^3 = e$. I will draw the corresponding neighborhood of ζ in the lecture.

Proof idea for Lemma 28.6. One checks that each $\sigma_0 \cdots \sigma_i \overline{P}$ has A as a vertex and angle $\angle_{\mathrm{int}} A_i$ at A where A_i is the initial vertex of s_i. When these angles sum to 2π, we get the result. □

Theorem 28.8. *If P is a locally finite fundamental domain for Γ, then Γ is generated by*

$$\{\gamma \in \Gamma \mid \gamma(\overline{P}) \cap \overline{P} \neq \varnothing\}.$$

If P is also convex, then Γ is generated by the side transformations σ_s (as s runs through the sides of P).

We defer the proof to [Ive92, V.2.3 and V.3.14].

Example 28.9. The modular group is generated by the side transformations $\{\sigma, \tau, \omega\}$ of M. Since $\omega = \tau^{-1}$, we can further cut the set of generators down to $\{\sigma, \tau\}$.

FIGURE 28. Johann Peter Gustav Lejeune Dirichlet (1805–59) was a German number theorist and analyst. In 1837, he proved that every arithmetic progression of integers contains infinitely many primes. He provided fundamental contributions to the theory of Fourier series, and is credited with the modern definition of a function ("to any x there corresponds a single y").

29. DIRICHLET DOMAINS

We now present Dirichlet's method for constructing fundamental domains. Again, $\Gamma \leq \mathrm{PGL}_2(\mathbb{R})$ is discrete throughout. Recall that if $K \subseteq \mathbb{H}^2$ is compact, then $\sigma(K) \cap K = \varnothing$ for only finitely many $\sigma \in \Gamma$. This allows us to pick $w \in \mathbb{H}^2$ such that $\Gamma_w = \{e\}$, that is, if $\sigma \in \Gamma$ satisfies $\sigma(w) = w$, then $\sigma = e$. Indeed, one can use proper discontinuity of a continuous group action on a locally compact set to prove that the collection of points with nontrivial stabilizer is discrete (moral exercise). Regardless, fix such a w.

Given $\sigma \in \Gamma$, let $L_\sigma(w)$ denote the perpendicular bisector of $[w, \sigma(w)]$, and let $H_\sigma(W)$ denote the component of the complement of $\mathbb{H}^2 \smallsetminus L_\sigma(w)$ containing w. The *Dirichlet domain* with center w is

$$P(w) = \bigcap_{\sigma \in \Gamma \smallsetminus \{e\}} H_\sigma(w).$$

We may think of $P(w)$ as the collection of points in $\mathbb{H}^2$ that are closer to w than they are to any Γ-translate of w. In other words, $P(w)$ is the *Voronoi cell* of w in the orbit Γw.

Theorem 29.1. *For $w \in \mathbb{H}^2$ such that $\Gamma_w = \{e\}$, the Dirichlet domain $P(w)$ is a convex locally finite fundamental domain for Γ in $\mathbb{H}^2$.*

Proof. We first check that $P(w)$ is open. While each $H_\sigma(w)$ is open, it is not immediate that $P(w)$ is open since arbitrary intersections of open sets need not be open. Observe, though, that for any $r > 0$ and $D(w; r)$ the open disk of radius r centered at w,

$$P(w) \cap D(w; r) = \bigcap_{\sigma \in S} H_\sigma(w)$$

for $S = \{\sigma \in \Gamma \smallsetminus \{e\} \mid \sigma(w) \in D(w; 2r)\}$. Since S is finite, each $P(w) \cap D(w; r)$ is open, and it follows that $P)(w)$ is open.

We now check that $P(w)$ is a fundamental domain. For $\gamma \in \Gamma$, the reader may verify that $\gamma P(w) = P(\gamma w)$, so $P(w) \cap \gamma P(w) = \varnothing$ for $\gamma \in \Gamma \smallsetminus \{e\}$. To show that the closure $\overline{P}(w)$ of $P(w)$ meets every Γ-orbit, first note that

$$\overline{P}(w) = \bigcap_{\sigma \in \Gamma \smallsetminus \{e\}} \overline{H}_\sigma(w).^{16} \tag{29.2}$$

Given a Γ-orbit $\mathscr{O}$, pick $z \in \mathscr{O}$ with shortest possible distance from w. Then

$$d(z, w) \leq d(\sigma^{-1}(z), w) = d(z, \sigma(w))$$

for $\sigma \in \Gamma \smallsetminus \{e\}$. Thus $z \in \overline{H}_\sigma(w)$ for all such σ, so $z \in \overline{P}(w)$ by (29.2).

Convexity of $P(w)$ is clear since it is the intersection of half-spaces. For local finitude, see [Ive92, Proposition V.4.4]. □

We have already seen a Dirichlet domain, namely the modular domain M for $\Gamma = \mathrm{PSL}_2(\mathbb{Z})$ and $w = yi$, $y > 1$. This follows from [Ive92, Lemma V.4.5], which says that if $S \subseteq \Gamma \smallsetminus \{e\}$ and

$$P_S(w) = \bigcap_{\sigma \in S} H_\sigma(w)$$

meets each Γ-orbit in at most one point, then $P_S(w) = P(w)$. The set M was constructed in this fashion for $S = \left\{ \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} \right\}$.

30. Compact polygons

We will skip Chapter VI of [Ive92], moving straight to the discussion of Poincaré's theorem in Chapter VII. The missing content (such as the monodromy theorem) will be mentioned *en passant* as we develop this final theorem.

Via the Dirichlet domain construction, we have seen that every discrete subgroup $\Gamma \leq \mathrm{PGL}_2(\mathbb{R})$ has a convex locally finite fundamental domain P in $\mathbb{H}^2$. When Γ is finitely generated, P has finitely many sides and is a polygon in $\mathbb{H}^2$ equipped with a side pairing $s \mapsto *s$. Our present goal is to understand which hyperbolic polygons and side pairings give rise to discrete groups of isometries.

Let Δ be a hyperbolic polygon in $\mathbb{H}^2$ with side pairing $s \mapsto *s$ on the set of edges $\mathscr{E}$ of Δ. We additionally require that s and $*s$ have the same length and that $*(s^{-1}) = (*s)^{-1}$ where s^{-1} is s with the opposite orientation. For $s \in \mathscr{E}$, there is a unique isometry σ_s mapping $*s$ to s in such a way that the half-plane bounded by $*s$ containing Δ is mapped to the half-plane bounded by s but opposite Δ. The isometry σ_s is called the *side transformation* determined by s. It is straightforward to observe that

$$\sigma_{s^{-1}} = \sigma_s \qquad \text{and} \qquad \sigma_{*s} = \sigma_s^{-1} \tag{30.1}$$

for $s \in \mathscr{E}$.

Given an edge s of Δ, let $\downarrow s$ denote the edge with the same initial vertex as s but different terminal vertex. Define the *edge operator*

$$\begin{aligned} \Psi \colon \mathscr{E} &\longrightarrow \mathscr{E} \\ s &\longmapsto \downarrow *s. \end{aligned}$$

[16] The inclusion $\subseteq$ is immediate since intersections of closed sets are closed. For the opposite inclusion, consider $z \in \mathbb{H}^2$ belonging to $\overline{H}_\sigma(w)$ for all $\sigma \in \Gamma \smallsetminus \{e\}$. Then $[w, z] \in \overline{H}_\sigma(w)$ for all σ, whence $[w, z)$ is in $P(w)$, so $z \in \overline{P}(w)$.

The edge operator is a bijection (both $*$ and $\downarrow$ are involutions) and $\mathscr{E}$ is finite, so Ψ has finite order.[17] The cycles for Ψ on $\mathscr{E}$ are called *edge cycles*. The sequence of initial points of an edge cycle is called a *vertex cycle*. (Note that vertices may repeat in a vertex cycle.)

Lemma 30.2. Let $s_1, \dots, s_r$ be an edge cycle with vertex cycle $P_1, \dots, P_r$ and side transformations $\sigma_i = \sigma_{s_i}$ for $1 \le i \le r$. The *cycle map* $\sigma = \sigma_1 \cdots \sigma_r$ is a rotation around P_1 of angle congruent modulo 2π to the sum of the interior angles

$$\angle_{\text{int}} P_1 + \angle_{\text{int}} P_2 + \cdots + \angle_{\text{int}} P_r.$$

Proof. Choose an orientation of $\mathbb{H}^2$ such that Δ is on the positive side of s_1 near P_1. Then

$$\angle_{\text{or}}(s_i, \downarrow s_i) \equiv \text{sign}(s_i) \angle_{\text{int}} P_i \pmod{2\pi}$$

for $1 \le i \le r$ where $\text{sign}(s_i)$ is $+1$ (resp. -1) if Δ lies on the positive (resp. negative) side of s_i. By the construction of the side transformations, $\text{sign}(s_m) = \text{sign}(s_{m+1}) \det(\sigma_m)$ for $1 \le m \le r$. Multiplying these equations together (and canceling $\text{sign}(s_2) \cdots \text{sign}(s_r)$) we get

$$\text{sign}(s_1) = \det(\sigma_1 \cdots \sigma_r) \, \text{sign}(s_{r+1}).$$

Since $s_{r+1} = s_1$, we conclude that the cycle map $\sigma_1 \cdots \sigma_r$ is an even isometry with fixed point P_1. We now claim that

$$\sum_{i=m}^{r} \angle_{\text{int}} P_i \equiv \text{sign}(s_m) \angle_{\text{or}}(s_m, \sigma_m \cdots \sigma_r(s_1))$$

for $1 \le m \le r$. To prove as much, observe that $s_i = \sigma_i(*s_i)$, $s_{i+1} = \downarrow * s_i$, and $\downarrow s_i = \sigma_i(s_{i+1})$. We now proceed by downward induction on m. When $m = r$, we have

$$\angle_{\text{int}} P_r = \text{sign}(s_r) \angle_{\text{or}}(s_r, \downarrow s_r) = \text{sign}(s_r) \angle_{\text{or}}(s_r, \sigma_r(s_1))$$

as desired. To descend from the $(m+1)$-th case to the m-th, observe that

$$\begin{aligned} \sum_{i=m}^{r} \angle_{\text{int}} P_i &\equiv \text{sign}(s_m) \angle_{\text{or}}(s_m, \downarrow s_m) + \text{sign}(s_{m+1}) \angle \text{or}(s_{m+1}, \sigma_{m+1} \cdots \sigma_r(s_1)) \\ &\equiv \text{sign}(s_m) \angle_{\text{or}}(s_m, \downarrow s_m) + \text{sign}(s_{m+1}) \det(\sigma_m) \angle_{\text{or}}(\downarrow s_m, \sigma_m \cdots \sigma_r(s_1)) \\ &\equiv \text{sign}(s_m) \angle_{\text{or}}(s_m, \downarrow s_m) + \text{sign}(s_m) \angle_{\text{or}}(\downarrow s_m, \sigma_m \cdots \sigma_r(s_1)) \\ &\equiv \text{sign}(s_m) \angle_{\text{or}}(s_m, \sigma_m \cdots \sigma_r(s_1)). \end{aligned}$$

The $m = 1$ case along with $\text{sign}(s_1) = 1$ gives the desired formula for the rotation angle. □

Example 30.3. Consider the hyperbolic octagon with side pairings pictured in Figure 29. Write $\alpha = \sigma_a$, $\beta = \sigma_b$, $\gamma = \sigma_c$, and $\delta = \sigma_d$. The edge cycle starting with a is

$$a(*b^{-1})(*a^{-1})bc(*d^{-1})(*c^{-1})d$$

with corresponding vertex cycle

$$P_1 P_4 P_3 P_2 P_5 P_8 P_7 P_6$$

and cycle map

$$\alpha\beta^{-1}\alpha^{-1}\beta\gamma\delta^{-1}\gamma^{-1}\delta.$$

[17] This means there is a positive integer n such that the n-fold composition $\Psi^n = \text{id}$; the smallest such n is the *order* of Ψ.

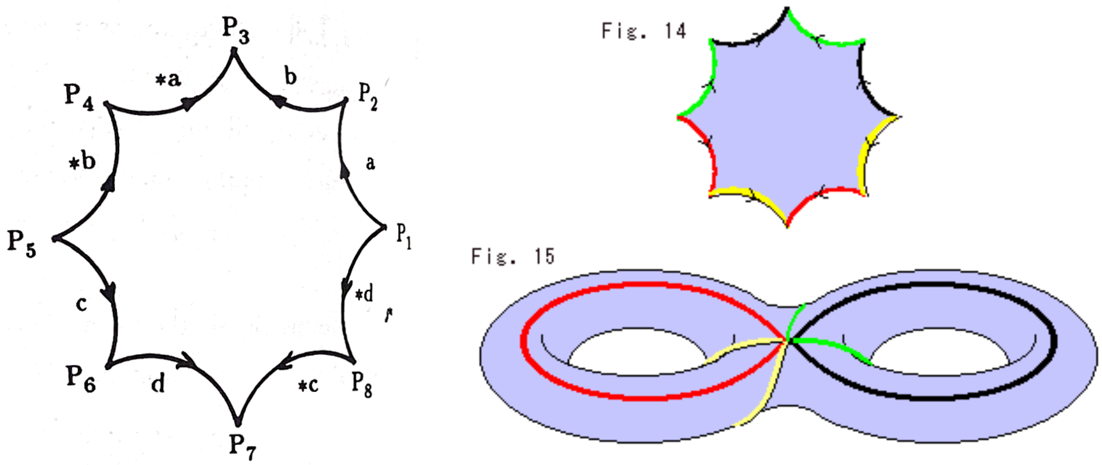

FIGURE 29. A hyperbolic octagon with indicated side pairings results in a genus 2 surface. Sources: [Ive92] and http://web1.kcn.jp/hp28ah77/us33_revo.htm.

31. POINCARÉ'S THEOREM

We now introduce the cycle condition and prove Poincaré's theorem on which convex polygons and side pairings generate discrete subgroups of $\mathrm{PGL}_2(\mathbb{R})$.

Definition 31.1. Fix a compact convex hyperbolic polygon Δ with side pairing $s \mapsto *s$. We say that an edge cycle $s_1, \dots, s_r$ for Δ with vertex cycle $P_1, \dots, P_r$ satisfies the *cycle condition* if the interior angle sum at the P_i takes the form

$$\sum_{i=1}^{r} \angle_{\mathrm{int}} P_i = \frac{2\pi}{n_c}$$

for some positive integer n_c. If all edge cycles of Δ satisfy the cycle condition, then we say that its side pairing satisfies the cycle condition.

By Lemma 30.2, the cycle map $\sigma = \sigma_1 \cdots \sigma_r$ of an edge cycle satisfying the cycle condition is a rotation about P_1 of order n_c.

Theorem 31.2 (Poincaré's theorem). *Let Δ be a compact convex hyperbolic polygon with a side pairing satisfying the cycle condition. Then the group Γ generated by the side transformations $S = \{\sigma_s \mid s \in \mathscr{E}\}$ is a discrete subgroup of $\mathrm{PGL}_2(\mathbb{R})$ with Δ as a fundamental domain. A presentation for Γ is given by $\langle S \mid R \rangle$ where R consists of the side relations (30.1) and the cycle relations*

$$\sigma_c^{n_c} = e \tag{31.3}$$

for each edge cycle c.

Example 31.4. Before diving into the proof of Theorem 31.2, let's return to Example 30.3 and consider the consequences. Suppose that the octagon presented there is regular with interior angle $\pi/4$.[18] Then there is a unique edge cycle with cycle map which is a rotation by angle $8 \cdot \pi/4 = 2\pi$, so $n_c = 1$ and $\sigma_c = e$. By Poincaré's theorem, the group generated by the side transformations of such an octagon is a discrete subgroup of $\mathrm{PGL}_2(\mathbb{R})$ with presentation

$$\left\langle \alpha, \beta, \gamma, \delta \mid \alpha\beta^{-1}\alpha^{-1}\beta\gamma\delta^{-1}\gamma^{-1}\delta \right\rangle.$$

[18] *Moral exercise*: Construct such an octagon.

If the octagon is regular with interior angle $\pi/(4n)$ for some positive integer n, then the group generated by the side transformations is again discrete but now with presentation

$$\left\langle \alpha, \beta, \gamma, \delta \mid (\alpha\beta^{-1}\alpha^{-1}\beta\gamma\delta^{-1}\gamma^{-1}\delta)^n \right\rangle .$$

Proof sketch for Theorem 31.2. Let $G = \langle S \mid R \rangle$. By Equation 30.1 and Lemma 30.2 we get a canonical homomorphism $G \to \Gamma$ denoted $g \mapsto \tilde{g}$. Give G the discrete topology and define a topological space

$$X = G \times \Delta / \sim$$

where $\sim$ is the equivalence relation generated by

$$(g\sigma_s, p) \sim (g, \sigma_s(p))$$

for $g \in G$, $s \in \mathscr{E}$, and $p \in *s$. Here $G \times \Delta$ has the product topology (open sets are of the form $\bigcup_{g \in G} \{g\} \times U_g$ where U_g is an open subset of Δ) and X has the quotient topology (open subsets correspond to open, $\sim$-stable subset of $G \times \Delta$). Write $g \cdot p \in X$ for the equivalence class of $(g, p) \in G \times \Delta$.

The evaluation map $G \times \Delta \to \mathbb{H}^2$, $(g, p) \mapsto \tilde{g}(p)$ is continuous and compatible with $\sim$, so it induces a continuous map $f\colon X \to \mathbb{H}^2$ where $f(g \cdot p) = \tilde{g}(p)$. Note that the action of G on the left-hand factor of $G \times \Delta$ induces an action of G on X (where $g(h \cdot p) = (gh) \cdot p$ for $g, h \in G$ and $p \in X$), and the map $f\colon X \to \mathbb{H}^2$ is equivariant in the sense that:

$$f(gx) = \tilde{g} f(x)$$

for $g \in G$ and $x \in X$.

At this point, we see that we have a "nice" map $f\colon X \to \mathbb{H}^2$. The proof proceeds by showing that f is in fact "very nice": X is a complete hyperbolic surface, and f is a local isometry. One then applies the monodromy theorem (our sin of omission, stated below as Theorem 31.7) to conclude that f is a bijection. Since f is surjective (and clearly $G \to \Gamma$ is surjective), we know that

$$\mathbb{H}^2 = \bigcup_{\gamma \in \Gamma} \gamma(\Delta).$$

By the definition of $\sim$, we have $g \cdot \Delta \cap e \cdot \Delta^\circ = \varnothing$ for $g \in G \smallsetminus \{e\}$. Since f is injective, it transforms disjoint sets into disjoint sets, and we learn that $\tilde{g}(\Delta) \cap \Delta^\circ = \varnothing$ for $g \in G \smallsetminus \{e\}$. This implies that $\ker(G \to \Gamma) = \{e\}$. (Indeed, if $\tilde{g} = e$ for some $g \in G \smallsetminus \{e\}$, then we would not have $\tilde{g}(\Delta)$ and Δ° disjoint!) Thus $G \to \Gamma$ is an isomorphism and Δ is a fundamental domain for Γ. □

In order to state the monodromy theorem, we need two definitions.

Definition 31.5. A *hyperbolic surface* is a metric space X with the shortest length property,[19] for which every point has an open neighborhood isometric to an open disk in $\mathbb{H}^2$. A hyperbolic surface is *complete* if every Cauchy sequence in the surface converges (in the surface).

Definition 31.6. A *local isometry* $f\colon X \to Y$ between metric spaces is a function such that every point in $x \in X$ has an open neighborhood which is mapped isometrically onto an open neighborhood of $f(x)$.

Theorem 31.7 (Monodromy theorem, see [Ive92, VI.4.8])**.** *Let X be a complete hyperbolic surface. Any local isometry $X \to \mathbb{H}^2$ is a bijection.*

The monodromy theorem has the following important corollary, which classifies complete hyperbolic surfaces.

[19] This property says that any two points x, y are joined by a geodesic of length $d(x, y)$. See [Ive92, Definition VI.1.2]

Corollary 31.8 ([Ive92, Theorem VI.6.1]). Any complete hyperbolic surface X is isometric to a surface of the form $\mathbb{H}^2/\Gamma$ where Γ is a torsion free discrete subgroup of $\mathrm{PGL}_2(\mathbb{R})$. Two such subgroups Γ and Σ define isometric surfaces $\mathbb{H}^2/\Gamma$ and $\mathbb{H}^2/\Sigma$ if and only if Γ and Σ are conjugate subgroups of $\mathrm{PGL}_2(\mathbb{R})$.

Example 31.9. We conclude with one more example of Poincaré's theorem in action. Let $\square ABCD$ be a convex quadrilateral in $\mathbb{H}^2$ where opposite sides have equal length. In your homework, you proved that the diagonals AC and BD have a common midpoint M and a half-turn with respect to M take the parallelogram to itself. We also know that the interior angle sum of $\square ABCD$ is less than 2π. Let us assume that this sum takes the form

$$\angle A + \angle B + \angle C + \angle D = \frac{2\pi}{n}$$

for some integer $n \geq 2$. Applying Theorem 31.2 to the side pairing $*AD = BC$ and $*AB = DC$ gives edge cycle

$$ab * a^{-1} * b^{-1}$$

and cycle map

$$\alpha\beta\alpha^{-1}\beta^{-1}$$

for $a = AD$, $b = BA$, $\alpha = \sigma_a$, and $\beta = \sigma_b$. Thus the parallelogram group Γ is isomorphic to

$$\left\langle \alpha, \beta \mid (\alpha\beta\alpha^{-1}\beta^{-1})^n \right\rangle.$$

32. Triangle groups

32.1. **Triangles.** We can use Poincaré's Theorem 31.2 to identify the discrete groups of isometries generated by hyperbolic triangles; these are the so-called *triangle groups*. Let $\triangle ABC$ be a triangle in $\mathbb{H}^2$. Let α denote reflection in BC, β reflection in CA, and γ reflection in AB. We assume that the interior angle measures of the triangle are

$$\angle A = \frac{\pi}{p}, \qquad \angle B = \frac{\pi}{q}, \qquad \angle C = \frac{\pi}{r}$$

for p, q, r positive integers such that

$$\frac{1}{p} + \frac{1}{q} + \frac{1}{r} < 1.$$

(The inequality guarantees that the total angle sum is less then π.) It turns out that for any three such positive integers, such a triangle exists and and is unique up to isometry. The subgroup of $\mathrm{Isom}(\mathbb{H}^2)$ generated by α, β, γ is denoted $\Delta(p, q, r)$ and is called the (p, q, r) *triangle group*.

The isometry $\gamma\beta$ is a rotation around A by angle $2\pi/p$, and similar remarks apply to vertices B and C. We thus have reflection and cycle relations amongst α, β, γ given by

$$e = \alpha^2 = \beta^2 = \gamma^2$$

and

$$e = (\gamma\beta)^p = (\alpha\gamma)^q = (\alpha\beta)^r.$$

We can apply Poincaré's theorem to the trivial pairing $*a = a, *b = b, *c = c$ with edge cycles $cb^{-1}, ac^{-1}, a^{-1}b$. This realizes the above relations, so

$$\Delta(p, q, r) \cong \left\langle \alpha, \beta, \gamma \mid \alpha^2, \beta^2, \gamma^2, (\gamma\beta)^p, (\alpha\gamma)^q, (\alpha\beta)^r \right\rangle$$

Now consider the even part of $\Delta(p, q, r)$,

$$D(p, q, r) := \Delta(p, q, r) \cap \mathrm{Isom}^+(\mathbb{H}^2).$$

FIGURE 30. Walther von Dyck (1856–1934) was a German mathematician and student of Felix Klein. He is credited with the modern definition of a mathematical group.

This is called the (p, q, r) *von Dyck group*. The reader may check that $D(p, q, r)$ is generated by $\rho_{2A} = \beta\gamma$, $\rho_{2B} = \gamma\alpha$, and $\rho_{2C} = \alpha\beta$ and we may derive the relations

$$\rho_{2A}\rho_{2B}\rho_{2C} = e = \rho_{2A}^p = \rho_{2B}^q = \rho_{2C}^r.$$

One can show that these are a complete list of relations for $D(p, q, r)$ by considering the quadrilateral $\square ABCD$ where D is the reflection of B across the geodesic AC. See [Ive92, p.210] for details.

32.2. **Hyperbolic pants.** Fix an orientation of the hyperbolic plane. Given a hyperbolic n-gon Δ, equip it with an oriented boundary cycle of edges $a_1, \ldots, a_n$ ($a_{n+1} = a_1$) where the terminal vertex of a_i matches the initial vertex of a_{i+1} and locally Δ is on the positive side of the edges of the boundary.

Let R be a rectangular hexagon in $\mathbb{H}^2$ with boundary cycle $a_1, \ldots, a_6$ and side pairing $*a_i = a_i$. By Poincaré's theorem, the reflections $\rho_1, \ldots, \rho_6$ in the sides of R generate a discrete group Π with fundamental domain R. The complete set of relations amongst the ρ_i is

$$\rho_i^2 = e \qquad \text{and} \rho_i\rho_{i+1} = \rho_{i+1}\rho_i$$

for $i = 1, \ldots, 6$ and $\rho_7 = \rho_1$.

Now consider a second copy of R and sew it to the first copy along a_2, a_4, a_6 as in Figure 31. This results in a pair of *hyperbolic pants* which are the orbit space of the rectangular octagon $\Delta = R \cup \rho_4(R)$ with side pairing indicated in the picture. The boundary cycle is $abc * b^{-1}def * e^{-1}$ and the side transformations are $\sigma_b = \rho_2\rho_4$, $\sigma_e = \rho_4\rho_6$. Applying Poincaré's theorem, we get that the group Γ generated by

$$\rho_1, \rho_3, \rho_5, \rho_{64}, \rho_{42}, \rho_{26}$$

(where $\rho_{ij} = \rho_i\rho_j$) is a discrete subgroup of index 2 in Π with fundamental domain Δ. The cycle and reflection relations may be written in the form

$$\rho_1^2 = \rho_3^2 = \rho_5^2 = \rho_{64}\rho_{42}\rho_{26} = e$$

$$\rho_1\rho_{26} = \rho_{26}\rho_1, \quad \rho_3\rho_{42} = \rho_{42}\rho_3, \quad \rho_5\rho_{64} = \rho_{64}\rho_5.$$

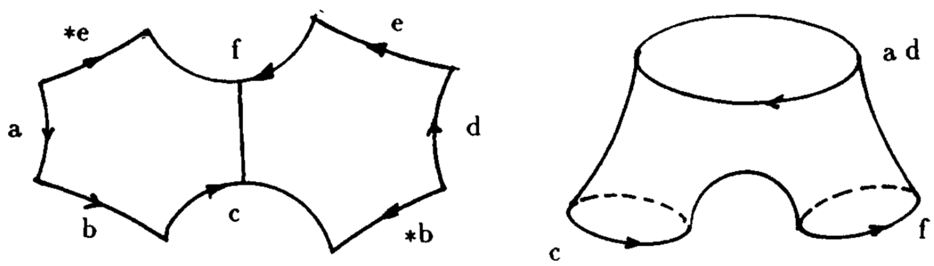

FIGURE 31. Hyperbolic pants assembled from two rectangular hexagons. Source: [Ive92, p.211].

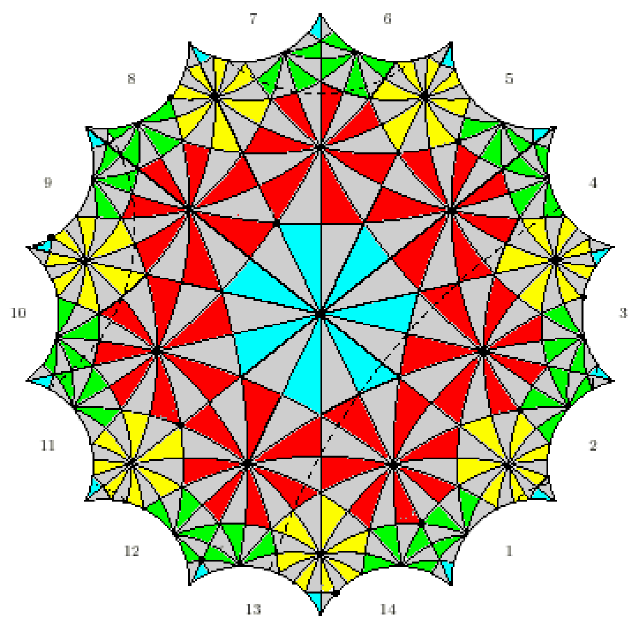

FIGURE 32. A fundamental domain of $\Gamma(7)$ in the Poincaré disk. Source: Tony Smith (via John Baez).

While we will not rehearse the details, our work with rectangular hexagons implies that hyperbolic surfaces homeomorphic to a pair of pants are uniquely determined (up to isometry) by the lengths $\ell_1, \ell_2, \ell_3 \in \mathbb{R}_{>0}$ of the the boundary components (cuffs and waist). This is important in the Fenchel–Nielsen coordinatization of Teichmüller space, but we will not pursue this further.

33. THE KLEIN QUARTIC

The *Klein quartic* is the surface $\mathbb{H}^2/\Gamma(7)$ where $\Gamma(7)$ is the kernel of the reduction map $\mathrm{PSL}_2(\mathbb{Z}) \to \mathrm{PSL}_2(\mathbb{F}_7)$. The fundamental domain of $\Gamma(7)$ is tiled by 24 heptagons, each of which is barycentrically subdivided into $(2,3,7)$ triangles. These regions and the side pairing induced by $\Gamma(7)$ are indicated in Figure 32. The resulting surface $\mathbb{H}^2/\Gamma(7)$ has genus 3 and automorphism group $\mathrm{PSL}_2(\mathbb{F}_7)$, which has order 168. It turns out that this is the maximal symmetry group of a genus 3 surface, making the Klein quartic a *Hurwitz surface*.

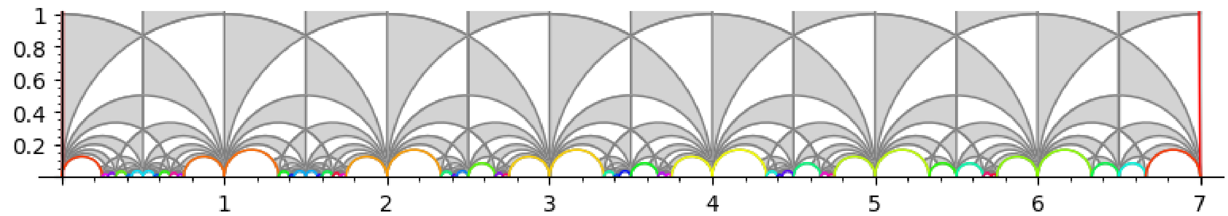

FIGURE 33. A fundamental domain of $\Gamma(7)$ in the upper half-plane produced by Sage.

John Baez maintains a delightful website devoted to the Klein quartic:

http://math.ucr.edu/home/baez/klein.html.

Please read it in preparation for our final class discussion.

REFERENCES

[Con] K. Conrad. Decomposing $\mathrm{SL}_2(\mathbb{R})$. Accessed March 2020 at https://kconrad.math.uconn.edu/blurbs/grouptheory/SL(2,R).pdf.

[Ive92] B. Iversen. *Hyperbolic geometry*, volume 25 of *London Mathematical Society Student Texts*. Cambridge University Press, Cambridge, 1992.

[Man15] K. Mann. DIY hyperbolic geometry. Accessed January 2020 at https://math.berkeley.edu/~kpmann/DIYhyp.pdf, 2015.

[Thu97] W.P. Thurston. *Three-dimensional geometry and topology. Vol. 1*, volume 35 of *Princeton Mathematical Series*. Princeton University Press, Princeton, NJ, 1997. Edited by Silvio Levy.