Chapter 27

Hyperbolic geometry

Having moved from algebra and arithmetic to analysis, in the final part of this text we consider geometric aspects of quaternion algebras: specifically, we consider spaces obtained from the unit group of a quaternion order.

We begin in the first chapter of this part with a rapid introduction to hyperbolic geometry. Hyperbolic geometry has its roots in efforts during the early 1800s to understand Euclid's axioms for geometry. Since the time of Euclid, there had been attempts to prove the quite puzzling parallel postulate (given a line and a point not on the line, there is a unique line through the point parallel to the given line) from the other four simpler, self-evident axioms for geometry. In hyperbolic geometry (a term coined by Klein, as its formulae can be obtained from those in spherical geometry by replacing trigonometric functions by their hyperbolic counterparts), the parallel postulate fails to hold—there are always infinitely many distinct lines through a point that do not intersect a given line—and so it is a non-Euclidean geometry. In this way, hyperbolic geometry is similar to Euclidean geometry and in some ways it is quite different: for example, there are no similarities in hyperbolic geometry—one cannot scale a figure without changing its shape and angles.

From our point of view, hyperbolic geometry is the natural place to study the unit group of a quaternion order. Indeed, it was work on automorphic functions invariant under discrete groups that provided Poincaré's original motivation for defining hyperbolic space: these discrete groups act by isometries on hyperbolic space and their study provides an incredibly rich interplay between number theory, algebra, geometry, and topology.

27.1 Geodesic spaces

In geometry, we need notions of length, distance, angle, and the straightness of a path. These notions are defined for a certain kind of metric space, as follows.

Let X be a metric space with distance ρ. An *isometry* $g : X \xrightarrow{\sim} X$ is a bijective map that preserves distance, i.e., $\rho(x, y) = \rho(g(x), g(y))$ for all $x, y \in X$. (Any distance-preserving map is automatically injective and so becomes an isometry onto its image.) The set of isometries $\mathrm{Isom}(X)$ naturally forms a group under composition.

27.1.1. A *path* from x to y, denoted $\gamma : x \to y$, is a continuous map $\gamma : [0, 1] \to X$ where $\gamma(0) = x$ and $\gamma(1) = y$. (More generally, we can take the domain to be any compact real interval.) The *length* $\ell(\gamma)$ of a path γ is the supremum of sums of distances between successive points over all finite subdivisions of the path (the path is *rectifiable* if this supremum is finite). Conversely, if X is a set with a notion of (nonnegative) length of path, then one recovers a candidate metric as

$$\rho(x, y) = \inf_{\gamma : x \to y} \ell(\gamma). \tag{27.1.2}$$

If the distance on X is of the form (27.1.2), we call X a *length space* or a *path metric space*, and by construction we have

$$\ell(g\gamma) = \rho(gx, gy) = \rho(x, y) = \ell(\gamma)$$

for all paths $\gamma : x \to y$ and $g \in \mathrm{Isom}(X)$.

Example 27.1.3. The space $X = \mathbb{R}^n$ with the ordinary Euclidean metric is a path metric space.

If X is a path metric space and γ achieves the infimum in (27.1.2), then we say γ is a *geodesic segment* in X. A *geodesic* is a continuous map $(-\infty, \infty) \to X$ such that the restriction to any compact interval defines a geodesic segment. If X is a path metric space such that any two points in X are joined by a geodesic segment, we say X is a *geodesic space*.

27.1.4. If X is a geodesic space, then an isometry of X maps geodesic segments to geodesic segments, and hence geodesics to geodesics: i.e., if $g \in \mathrm{Isom}(X)$ and $\gamma : x \to y$ is a geodesic segment, then $(g\gamma) : gx \to gy$ is a geodesic sesgment. After all, g maps the set of paths $x \to y$ bijectively to the set of paths $gx \to gy$ bijectively, preserving distance.

27.1.5. In the context of differential geometry (our primary concern), these notions can be made concrete enough to do computations. Suppose $U \subseteq \mathbb{R}^n$ is an open subset. Then a convenient way to specify the length of a path in U is with a *length element* in real-valued coordinates. To illustrate, the ordinary metric on $\mathbb{R}^n$ is given by the length element

$$\mathrm{d}s = \sqrt{\mathrm{d}x_1^2 + \cdots + \mathrm{d}x_n^2},$$

so if $\gamma : [0, 1] \to U$ is piecewise continuously differentiable, then

$$\ell(\gamma) = \int_\gamma \sqrt{(\mathrm{d}x_1/\mathrm{d}t)^2 + \cdots + (\mathrm{d}x_n/\mathrm{d}t)^2}\ \mathrm{d}t \tag{27.1.6}$$

as usual. More generally, if $\lambda : U \to \mathbb{R}_{>0}$ is a positive continuous function, then the length element $\lambda(x)\,\mathrm{d}s$ defines a metric (27.1.2) on U: the associated length (27.1.6) is symmetric, nonnegative, satisfies the triangle inequality, and to show that $\rho(x, y) > 0$ when $x \neq y$, by continuity λ is bounded below by some $\eta > 0$ on a suitably small ϵ ball neighborhood of x not containing y, so any path $\gamma : x \to y$ has $\ell(\gamma) \geq \epsilon\eta$ so $\rho(x, y) > 0$. In this context, we also have a notion of orientation, and we may ask that for isometries that preserve this orientation.

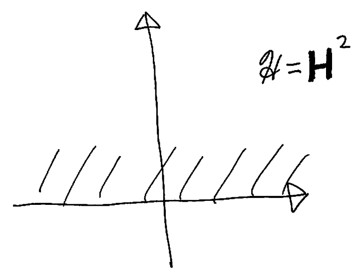
We return to this in section 27.8, rephrasing it terms of Riemannian geometry.

27.2 Upper half-plane

We now present the first model of two-dimensional hyperbolic space.

Definition 27.2.1. The *upper half-plane* is the set

$$\mathbf{H}^2 = \{z = x + iy \in \mathbb{C} : \mathrm{Im}(z) = y > 0\}.$$



For uniformity with some formulas that will appear later and for conciseness of formulas, we will write simply $\mathcal{H} = \mathbf{H}^2$.

Definition 27.2.2. The *hyperbolic length element* on $\mathcal{H}$ is defined by

$$\mathrm{d}s = \frac{\sqrt{\mathrm{d}x^2 + \mathrm{d}y^2}}{y} = \frac{|\mathrm{d}z|}{\mathrm{Im}\, z}; \tag{27.2.3}$$

As described in Paragraph 27.1.5, the hyperbolic length element induces a metric on $\mathcal{H}$, and this provides it with the structure of a path metric space.

Definition 27.2.4. The set $\mathbf{H}^2$ equipped with the hyperbolic metric is (a model for) the *hyperbolic plane*.

Remark 27.2.5. The space $\mathbf{H}^2$ can be intrinsically characterized as the unique two-dimensional (connected and) simply connected Riemannian manifold with constant sectional curvature -1. See section 27.9 for further reference.

The hyperbolic metric and the Euclidean metric on $\mathcal{H}$ are equivalent, inducing the same topology (Exercise 27.1). However, lengths and geodesics are different, as we will soon see.

27.2.6. The group $\mathrm{GL}_2(\mathbb{R})$ acts on $\mathbb{C}$ via linear fractional transformations:

$$gz = \begin{pmatrix} a & b \\ c & d \end{pmatrix} z = \frac{az+b}{cz+d}, \qquad \text{for } g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{R}) \text{ and } z \in \mathbb{C};$$

since

$$gz = \frac{(az+b)(c\overline{z}+d)}{|cz+d|^2} = \frac{ac|z|^2 + adz + bc\overline{z} + bd}{|cz+d|^2}$$

we have

$$\mathrm{Im}\, gz = \frac{\det g}{|cz+d|^2}\, \mathrm{Im}\, z. \tag{27.2.7}$$

and so $\operatorname{Im} z > 0$ if and only if $\operatorname{Im} gz > 0$. Therefore, the subgroup

$$\mathrm{GL}_2^+(\mathbb{R}) = \{g \in \mathrm{GL}_2(\mathbb{R}) : \det(g) > 0\}$$

preserves the upper half-plane $\mathcal{H}$. Moreover, because the action of $\mathrm{GL}_2^+(\mathbb{R})$ is holomorphic, it is orientation-preserving.

Scalar matrices act by the identity as linear fractional transformations, so taking the quotient we get a faithful action of $\mathrm{PGL}_2^+(\mathbb{R}) = \mathrm{GL}_2^+(\mathbb{R})/\mathbb{R}^\times$ on $\mathcal{H}$. We have a canonical isomorphism

$$\begin{aligned} \mathrm{PGL}_2^+(\mathbb{R}) &\xrightarrow{\sim} \mathrm{PSL}_2(\mathbb{R}) \\ g &\mapsto \frac{1}{\sqrt{\det(g)}} g \end{aligned}$$

with the same action on the upper half-plane.

27.2.8. The determinant $\det : \mathrm{PGL}_2(\mathbb{R}) \to \mathbb{R}^\times/\mathbb{R}^{\times 2} \simeq \{\pm 1\}$ has the inverse image of $+1$ equal to $\mathrm{PGL}_2^+(\mathbb{R})$ both open and closed in $\mathrm{PGL}_2(\mathbb{R})$; therefore, any g with $\det(g) < 0$ together with $\mathrm{PGL}_2^+(\mathbb{R})$ generates $\mathrm{PGL}_2(\mathbb{R})$: for example, we may take

$$g = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \tag{27.2.9}$$

By (27.2.7), therefore, we extend the action of $\mathrm{PGL}_2(\mathbb{R})$ on $\mathcal{H}$ by defining for $g \in \mathrm{PGL}_2(\mathbb{R})$ and $z \in \mathcal{H}$

$$gz = \begin{cases} gz, & \text{if } \det g > 0; \\ g\overline{z}, & \text{if } \det g < 0. \end{cases} \tag{27.2.10}$$

The elements $g \in \mathrm{PGL}_2(\mathbb{R})$ with $\det g < 0$ act anti-holomorphically and so are orientation-reversing. The matrix g in (27.2.9) then acts by $g(z) = -\overline{z}$

This action also arises naturally from another point of view. Let

$$\mathcal{H}^- = \{z \in \mathbb{C} : \operatorname{Im} z < 0\}$$

be the lower half-plane, let $\mathcal{H}^+ = \mathcal{H}$, and let

$$\mathcal{H}^\pm = \mathcal{H}^+ \cup \mathcal{H}^- = \{z \in \mathbb{C} : \operatorname{Im} z \neq 0\} = \mathbb{C} \setminus \mathbb{R}.$$

Then $\mathrm{PGL}_2(\mathbb{R})$ acts on $\mathcal{H}^\pm$ (it preserves $\mathbb{R}$ so too the complement). Complex conjugation $z \mapsto \overline{z}$ interchanges $\mathcal{H}^+$ and $\mathcal{H}^-$, so we have a canonical identification

$$\mathcal{H}^\pm/\langle \overline{} \rangle \xrightarrow{\sim} \mathcal{H}$$

and therefore obtain the action (27.2.10) of $\mathrm{PGL}_2(\mathbb{R})$ on $\mathcal{H}$.

Remark 27.2.11. The fact that $\mathrm{PSL}_2(\mathbb{R})$ has elements $g \in \mathrm{PSL}_2(\mathbb{R})$ that are matrices up to sign means that whenever we do a computation with a choice of matrix, implicitly we are also checking that the computation goes through with the other choice of sign. Most of the time, this is harmless—but in certain situations this sign plays an important role!

Theorem 27.2.12. *The group* $\mathrm{PSL}_2(\mathbb{R})$ *acts on* $\mathcal{H}$ *via orientation-preserving isometries, i.e.,* $\mathrm{PSL}_2(\mathbb{R}) \hookrightarrow \mathrm{Isom}^+(\mathcal{H})$.

Proof. Because the metric is defined by a length element $\mathrm{d}s$, we want to show that $\mathrm{d}(gs) = \mathrm{d}s$ for all $g \in \mathrm{PSL}_2(\mathbb{R})$, i.e.,

$$\frac{|\mathrm{d}(gz)|}{\mathrm{Im}(gz)} = \frac{|\mathrm{d}z|}{\mathrm{Im}\, z}$$

for all $g \in \mathrm{PSL}_2(\mathbb{R})$. Since $|\mathrm{d}(gz)| = |\mathrm{d}g/\mathrm{d}z||\mathrm{d}z|$, it is equivalent to show that

$$\frac{|\mathrm{d}g/\mathrm{d}z|}{\mathrm{Im}\, gz} = \frac{1}{\mathrm{Im}\, z} \tag{27.2.13}$$

for all $g \in \mathrm{PSL}_2(\mathbb{R})$.

Let $g \in \mathrm{PSL}_2(\mathbb{R})$ act by

$$g(z) = \frac{az+b}{cz+d}$$

with $ad - bc = 1$. Then

$$|\mathrm{d}g/\mathrm{d}z| = \left|\frac{(cz+d)a - (az+b)c}{(cz+d)^2}\right| = \frac{1}{|cz+d|^2}; \tag{27.2.14}$$

by (27.2.7), we have

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

so taking the ratio, the two factors $|cz+d|^2$ exactly cancel, establishing (27.2.13).

If the above calculus with metrics is a bit opaque, you can also work directly from the definitions. Let υ be a (piecewise continuously differentiable) path in $\mathcal{H}$ with $\upsilon : [0,1] \to \mathcal{H}$ given by $z(t)$; then by definition

$$\ell(\upsilon) = \int_0^1 \left|\frac{\mathrm{d}z}{\mathrm{d}t}\right| \frac{\mathrm{d}t}{\mathrm{Im}\, z(t)}.$$

The path $g\upsilon$ is given by $w(t) = g(z(t))$, so by the chain rule, we have

$$\ell(g(\upsilon)) = \int_0^1 \left|\frac{\mathrm{d}w}{\mathrm{d}t}\right| \frac{\mathrm{d}t}{\mathrm{Im}\, w(t)} = \int_0^1 \left|\frac{\mathrm{d}g}{\mathrm{d}z}\frac{\mathrm{d}z}{\mathrm{d}t}\right| \frac{\mathrm{d}t}{\mathrm{Im}\, g(z(t))} = \int_0^1 \left|\frac{\mathrm{d}z}{\mathrm{d}t}\right| \frac{\mathrm{d}t}{\mathrm{Im}\, z(t)},$$

the latter equality from (27.2.13). The fact that lengths are preserved immediately implies the invariance of the hyperbolic metric.

In any event, the action is holomorphic so (by the Cauchy–Riemann equations) lands in the orientation-preserving subgroup. $\square$

27.2.15. The action of $\mathrm{PSL}_2(\mathbb{R})$ extends to the boundary as follows. We define the *circle at infinity* to be the boundary

$$\partial\mathbf{H}^2 = \mathbb{R} \cup \{\infty\} \subseteq \mathbb{C} \cup \{\infty\}.$$

The $\mathrm{PSL}_2(\mathbb{R})$ acts on $\partial\mathsf{H}^2$. We define the *completed upper half-plane* to be

$$\mathsf{H}^{2*} = \mathsf{H}^2 \cup \partial\mathsf{H}^2.$$

The topology on H^{2*} is the same as the Euclidean topology on $\mathsf{H}^2 \cup \mathbb{R}$, and we take a fundamental system of neighborhoods of the point at ∞ to be sets of the form

$$\{z \in \mathsf{H}^2 : \operatorname{Im} z > M\} \cup \{\infty\}$$

for $M > 0$.

27.3 Classification of isometries

On our way to classifying isometries, we pause to identify three natural subgroups of $\mathrm{SL}_2(\mathbb{R})$:

$$\begin{aligned} K = \mathrm{SO}(2) &= \left\{ \begin{pmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{pmatrix} : t \in \mathbb{R} \right\} \simeq \mathbb{R}/\mathbb{Z} \\ A &= \left\{ \begin{pmatrix} a & 0 \\ 0 & 1/a \end{pmatrix} : a \in \mathbb{R}_{>0}^{\times} \right\} \simeq \mathbb{R}_{>0}^{\times} \simeq \mathbb{R} \\ N &= \left\{ \begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix} : n \in \mathbb{R} \right\} \simeq \mathbb{R}. \end{aligned} \tag{27.3.1}$$

We have $K = \mathrm{Stab}_{\mathrm{SL}_2(\mathbb{R})}(i)$ since $(ai+b)/(ci+d) = i$ if and only if $d = a$ and $c = -b$, and then the determinant condition implies $|a|, |b| \le 1$. An element $\begin{pmatrix} a & 0 \\ 0 & 1/a \end{pmatrix}$ acts by $z \mapsto a^2 z$, fixing the origin and stretching along lines through the origin. An element $\begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix}$ acts by $z \mapsto z + n$, which is translation.

Proposition 27.3.2 (Iwasawa decomposition)**.** *The multiplication map gives a homeomorphism*

$$N \times A \times K \xrightarrow{\sim} \mathrm{SL}_2(\mathbb{R}).$$

In particular, for all $g \in \mathrm{SL}_2(\mathbb{R})$, we can write uniquely $g = n_g a_g k_g$ with $n_g \in N$, $a_g \in A$, and $k_g \in K$ in a continuous way. This proposition together with (27.3.1) gives us another way to see that $\mathrm{SL}_2(\mathbb{R})$ is three-dimensional.

Proof. The multiplication map $N \times A \times K \to \mathrm{SL}_2(\mathbb{R})$ is continuous and open, so we need to show it is bijective. It is injective, because checking directly we see that

$$NA \cap K = \{1\} = N \cap A.$$

This map is surjective as follows. Let $g \in \mathrm{SL}_2(\mathbb{R})$, and let $z = g(i)$. Let $n_g = \begin{pmatrix} 1 & -\operatorname{Re} z \\ 0 & 1 \end{pmatrix} \in N$, so that $(n_g g)(i) = si$ with $s > 0$. Let $a_g = \begin{pmatrix} 1/\sqrt{s} & 0 \\ 0 & \sqrt{s} \end{pmatrix} \in A$; then $(a_g n_g g)(i) = i$, so $a_g n_g g \in \mathrm{Stab}_{\mathrm{SL}_2(\mathbb{R})}(i) \simeq \mathrm{SO}(2)$, and peeling back we get $g \in NAK$, proving surjectivity. $\square$

These subgroups can be characterized by their traces; with a view to working on $\mathrm{PSL}_2(\mathbb{R})$, we consider the absolute traces: we have

$$|\mathrm{Tr}(K)| = [0,2], \ |\mathrm{Tr}(A)| = [2,\infty), \text{ and } |\mathrm{Tr}(N)| = \{2\}.$$

Definition 27.3.3. An element $g \in \mathrm{PSL}_2(\mathbb{R})$ with $g \neq \pm 1$ is called *elliptic*, *hyperbolic*, or *parabolic* according to whether $|\mathrm{Tr}(g)| < 2$, $|\mathrm{Tr}(g)| > 2$, or $|\mathrm{Tr}(g)| = 2$.

Every nonidentity element $g \in \mathrm{PSL}_2(\mathbb{R})$ belongs to one of these three types.

Lemma 27.3.4. *An element $g \in \mathrm{PSL}_2(\mathbb{R})$ is elliptic if and only if g has a unique fixed point in $\mathbf{H}^2$, hyperbolic if and only if g has two fixed points on $\partial\mathbf{H}^2$, and parabolic if and only if g has a unique fixed point on $\partial\mathbf{H}^2$.*

Proof. Let $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ and $\det(g) = ad - bc = 1$. We look to solve the equation

$$\frac{az+b}{cz+d} = z$$

or equivalently

$$cz^2 + (d-a)z - b = 0.$$

The discriminant is

$$(d-a)^2 + 4bc = (a+d)^2 - 4 = \mathrm{Tr}(g)^2 - 4.$$

Therefore g is elliptic if and only if this discriminant is negative if and only if there is a unique root in $\mathcal{H}$; g is parabolic if and only if this discriminant is zero if and only if there is a unique root in $\partial\mathcal{H}$; and g is hyperbolic if and only if this discriminant is positive if and only if there are two roots in $\partial\mathcal{H}$. □

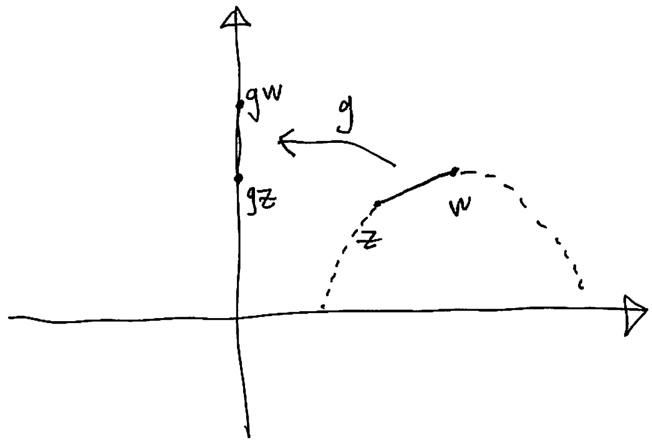
27.3.5. Let $g \in \mathrm{PSL}_2(\mathbb{R})$. If g is elliptic, then g acts by rotation around its fixed point; every such element is conjugate to an element of K, fixing i.

A hyperbolic element can be thought of as a translation along the geodesic between the two fixed points on $\partial\mathbf{H}^2$; every such element is conjugate to an element of A, acting by $z \mapsto a^2 z$ with $a \neq 1$.

Finally, a parabolic element should be thought of as a limit of the other two types, where correspondingly the fixed point tends to the boundary or the two fixed points move together; every such element is conjugate to an element of N, acting by translation $z \mapsto z + n$ for some $n \in \mathbb{R}$.

The following lemma can then be proved directly using translations and scaling (27.3.1).

Lemma 27.3.6. *For any two points $z, z' \in \mathcal{H}$, there exists $g \in \mathrm{PSL}_2(\mathbb{R})$ such that $gz, gz' \in \mathbb{R}_{>0} i$ are pure imaginary.*



Proof. Exercise 27.5. □

27.4 Geodesics

In this section, using the above we prove two important theorems: we describe geodesics, give a formula for the distance, and classify isometries.

Theorem 27.4.1. *The hyperbolic plane $\mathcal{H}$ is a geodesic space. In fact, there is a unique geodesic passing through any two points $z, z' \in \mathcal{H}$, a semicircle orthogonal to $\mathbb{R}$ or a vertical line, and*

$$\rho(z, z') = \log \frac{|z - \overline{z'}| + |z - z'|}{|z - \overline{z'}| - |z - z'|} \tag{27.4.2}$$

$$\cosh \rho(z, z') = 1 + \frac{|z - z'|^2}{2 \operatorname{Im}(z) \operatorname{Im}(z')}. \tag{27.4.3}$$

Proof. We first prove the imaginary axis is a geodesic with $z, z' \in \mathbb{R}_{>0} i$. Let $\upsilon(t) = x(t) + iy(t) : z \to z'$ be a path; then

$$\begin{aligned}\ell(\upsilon) &= \int_0^1 \frac{\sqrt{(\mathrm{d}x/\mathrm{d}t)^2 + (\mathrm{d}y/\mathrm{d}t)^2}}{y(t)} \,\mathrm{d}t \geq \int_0^1 \frac{\mathrm{d}y/\mathrm{d}t}{y(t)} \,\mathrm{d}t \\ &= \log y(1) - \log y(0) = \log \left| \frac{z}{z'} \right| \end{aligned} \tag{27.4.4}$$

with equality if and only if $x(t) = 0$ identically and $\mathrm{d}y/\,\mathrm{d}t \geq 0$ for all $t \in [0, 1]$. This is achieved for the path $\upsilon(t) = (|z|(1 - t) + |z'|t)i$; so $\rho(z, z') = \log|z/z'|$, and the imaginary axis is the unique geodesic.

For arbitrary points $z, z' \in \mathcal{H}$, we apply Lemma 27.3.6. The statement on geodesics follows from the fact that the image of $\mathbb{R}_{>0} i$ under an element of $\mathrm{PSL}_2(\mathbb{R})$ is either a semicircle orthogonal to $\mathbb{R}$ or a vertical line (Exercise 27.6).

The formula (27.4.2) for the case $z, z' \in \mathbb{R}_{>0}i$ follows from (27.4.4) and the equality

$$\frac{s}{s'} = \frac{(s+s')+(s-s')}{(s+s')-(s-s')}$$

for $s > s'$ in $\mathbb{R}$; the general case then follows from the invariance of both $\rho(z,z')$ and

$$\log \frac{|z-\overline{z'}|+|z-z'|}{|z-\overline{z'}|-|z-z'|}$$

under $g \in \mathrm{PSL}_2(\mathbb{R})$ (Exercise 27.8). Finally, the formula (27.4.3) follows from a direct computation, requested in Exercise 27.9. □

Theorem 27.4.5. *We have*

$$\mathrm{Isom}(\mathcal{H}) \simeq \mathrm{PGL}_2(\mathbb{R})$$

and

$$\mathrm{Isom}^+(\mathcal{H}) \simeq \mathrm{PGL}_2^+(\mathbb{R}) \simeq \mathrm{PSL}_2(\mathbb{R}).$$

Proof. Let $Z = \{ti : t > 0\}$ be the positive part of the imaginary axis. By Theorem 27.4.1, Z is the unique geodesic through any two points of Z.

Let $\phi \in \mathrm{Isom}(\mathcal{H})$. Then $\phi(Z)$ is a geodesic (27.1.4), so by Exercise 27.7, there exists $g \in \mathrm{PSL}_2(\mathbb{R})$ such that $g\phi$ fixes Z pointwise. Replacing ϕ by $g\phi$, we may assume ϕ fixes Z pointwise.

Let $z = x + iy \in \mathcal{H}$ and $z' = x' + iy' = \phi(z)$. For all $t > 0$, we have

$$\rho(z, it) = \rho(\phi z, \phi(it)) = \rho(z', it).$$

Plugging this into the formula (27.4.3) for the distance, we obtain

$$(x^2 + (y-t)^2)y' = (x'^2 + (y'-t)^2)y.$$

Dividing both sides by t^2 and taking the limit as $t \to \infty$, we find that $y = y'$, and consequently that $x^2 = x'^2$ so $x = \pm x'$. By continuity, the choice of sign $\pm$ various continuously over $\mathcal{H}$ and so must be constant. Therefore $\phi(z) = z$ or $\phi(z) = -\overline{z}$ for all $z \in \mathcal{H}$. The latter generates $\mathrm{PGL}_2(\mathbb{R})$ over $\mathrm{PSL}_2(\mathbb{R})$ (27.2.8), so both statements in the theorem follow. □

27.5 Hyperbolic area and the Gauss–Bonnet formula

We obtain an area form from the hyperbolic length element by considering a small Euclidean rectangle whose sides are parallel to the axes at the point (x, y); the hyperbolic length of the sides are $\mathrm{d}x/y$ and $\mathrm{d}y/y$, and we obtain the hyperbolic area form from the product.

Definition 27.5.1. We define the *hyperbolic area form* by

$$\mathrm{d}A = \frac{\mathrm{d}x\,\mathrm{d}y}{y^2}.$$

For a subset $S \subseteq \mathcal{H}$, we define the *hyperbolic area* of S by

$$\mu(S) = \iint_S \mathrm{d}A$$

when this integral is defined.

Proposition 27.5.2. *The hyperbolic area is invariant under* $\operatorname{Isom}(\mathcal{H})$.

Proof. We first check this for the orientation-reversing isometry

$$g(z) = g(x+iy) = -x+iy = -\overline{z};$$

visibly, we have $\mathrm{d}(gA) = \mathrm{d}A$ in this case.

It suffices then to consider $g \in \mathrm{PSL}_2(\mathbb{R})$. Let $z = x+iy$ and let

$$w = g(z) = \frac{az+b}{cz+d} = u+iv,$$

with $ad-bc=1$. By (27.2.7), we have $v = \dfrac{y}{|cz+d|^2}$. We compute that

$$\frac{\mathrm{d}g}{\mathrm{d}z} = \frac{1}{(cz+d)^2}. \tag{27.5.3}$$

Now g is holomorphic; so by the Cauchy–Riemann equations, its Jacobian is given by

$$\frac{\partial(u,v)}{\partial(x,y)} = u_x^2 + v_y^2 = |\mathrm{d}g/\mathrm{d}z|^2 = \frac{1}{|cz+d|^4};$$

therefore

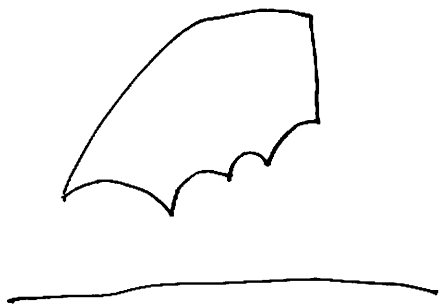
$$\mathrm{d}(gA) = \frac{\mathrm{d}u\,\mathrm{d}v}{v^2} = \frac{\partial(u,v)}{\partial(x,y)}\frac{\mathrm{d}x\,\mathrm{d}y}{v^2} = \frac{1}{|cz+d|^4}\frac{|cz+d|^4}{y^2}\,\mathrm{d}x\,\mathrm{d}y = \mathrm{d}A. \qquad \square$$

Definition 27.5.4. Let $z, z' \in \mathcal{H}^*$ be distinct points. Then there is a unique geodesic whose closure in $\mathcal{H}^*$ passes through z, z', called *sides*. A *hyperbolic polygon* is a connected, closed subset of $\mathcal{H}^*$ whose boundary consists of finitely many sides. A point of intersection between two sides is called a *vertex*.

A hyperbolic polygon P is *convex* if the geodesic segment between any two points in P lies inside P.

A *hyperbolic triangle* is a hyperbolic polygon with three sides.

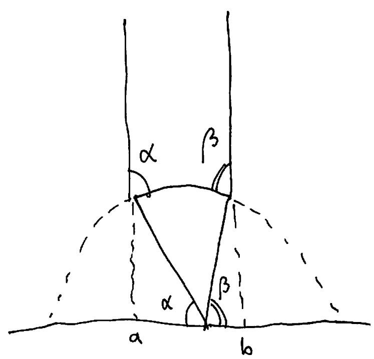
A hyperbolic triangle is necessarily convex, and so it can alternatively be described as the convex hull of its three vertices.



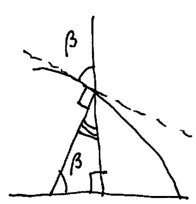
Theorem 27.5.5 (Gauss–Bonnet formula)**.** *Let T be a hyperbolic triangle with angles α, β, γ. Then*

$$\mu(T) = \pi - \alpha - \beta - \gamma.$$

Proof. We first consider the case where T has exactly one vertex in $\partial\mathcal{H}$. Applying an element of $\mathrm{PSL}_2(\mathbb{R})$, we may assume this vertex is $i\infty$. Then we have a diagram as follows:



The fact that the angles are duplicated along the real axis is explained by the following diagram:



The semicircular segment lies along a circle with some radius c; applying the isometry

$$g = \begin{pmatrix} 1/\sqrt{c} & 0 \\ 0 & \sqrt{c} \end{pmatrix} \in \mathrm{PSL}_2(\mathbb{R})$$

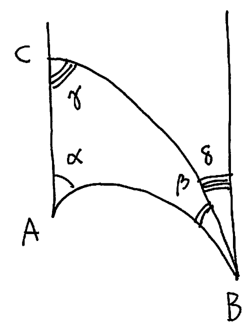
with the effect $g(z) = z/c$, we may assume that this segment lies along the unit circle. Then

$$\iint_T \frac{\mathrm{d}x\,\mathrm{d}y}{y^2} = \int_a^b \int_{\sqrt{1-x^2}}^{\infty} \frac{\mathrm{d}y}{y^2}\,\mathrm{d}x = \int_a^b \left.\frac{-1}{y}\right|_{\sqrt{1-x^2}}^{\infty} \mathrm{d}x$$
$$= \int_a^b \frac{\mathrm{d}x}{\sqrt{1-x^2}} = \int_{\pi-\alpha}^{\beta} -\,\mathrm{d}\theta = \pi - (\alpha - \beta)$$

where we make the substitution $x = \cos\theta$.

The case where T has two or three vertices in $\partial\mathcal{H}$ is left as an exercise (Exercise 27.10).

So we are left with the case where T has all vertices in $\mathcal{H}$. We then consider the following diagram:



We find that the area is

$$(\pi - (\alpha+\beta)) - (\pi - ((\pi-\delta)+\delta)) = \pi - (\alpha+\beta+\gamma)$$

as desired. $\square$

27.5.6. Let P be a convex hyperbolic polygon with n sides. By convexity, each side meets at each vertex a unique side, so P has n vertices with angles $\theta_1, \dots, \theta_n$. Triangulating P and applying the Gauss–Bonnet theorem, we conclude that

$$\mu(P) = (n-2)\pi - \sum_{n=1}^{\infty} \theta_n.$$

27.6 Unit disc and Lorentz model

In this section, we consider two other models for the hyperbolic plane.

First, we consider the unit disc model.

Definition 27.6.1. The *hyperbolic unit disc* is the (open) unit disc

$$\mathbf{D}^2 = \{w \in \mathbb{C} : |w| < 1\}$$

equipped with the *hyperbolic metric*

$$\mathrm{d}s = \frac{2|\mathrm{d}w|}{1-|w|^2}.$$

The hyperbolic unit disc $\mathbf{D}^2$ is also called the *Poincaré model* of planar hyperbolic geometry. We will also use the notation $\mathcal{D} = \mathbf{D}^2$ when convenient. The *circle at infinity* is the boundary

$$\partial\mathbf{D}^2 = \{w \in \mathbb{C} : |w| = 1\}.$$

27.6.2. For any $z_0 \in \mathbf{H}^2$, the maps

$$\begin{aligned} \phi : \mathbf{H}^2 &\xrightarrow{\sim} \mathbf{D}^2 & \phi^{-1} : \mathbf{D}^2 &\xrightarrow{\sim} \mathbf{H}^2 \\ z &\mapsto w = \frac{z-z_0}{z-\overline{z_0}} & w &\mapsto z = \frac{\overline{z_0}w - z_0}{w-1} \end{aligned} \tag{27.6.3}$$

define a conformal equivalence between $\mathbf{H}^2$ and $\mathbf{D}^2$ with $z_0 \mapsto \phi(z_0) = 0$. A particularly nice choice is $z_0 = i$, giving

$$\phi(z) = \frac{z-i}{z+i}, \quad \phi^{-1}(w) = -i\frac{w+1}{w-1}. \tag{27.6.4}$$

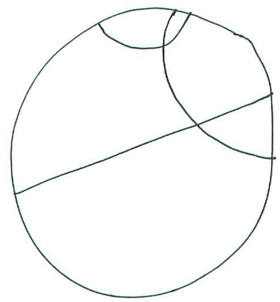
The hyperbolic metric on $\mathbf{D}^2$ is the pushforward of (induced from) the hyperbolic metric on $\mathbf{H}^2$ via the identification (27.7.15) (Exercise 27.11). Ordinarily, one would decorate the pushforward metric, but because we will frequently move between the upper half-plane and unit disc as each has its advantage, we find it notationally simpler to avoid this extra decoration. The distance on $\mathbf{D}^2$ is

$$\begin{aligned} \rho(w,w') &= \log\frac{|1-w\overline{w'}| + |w-w'|}{|1-w\overline{w'}| - |w-w'|} \\ \cosh\rho(w,w') &= 1 + 2\frac{|w-w'|^2}{(1-|w|^2)(1-|w'|^2)} \end{aligned}$$

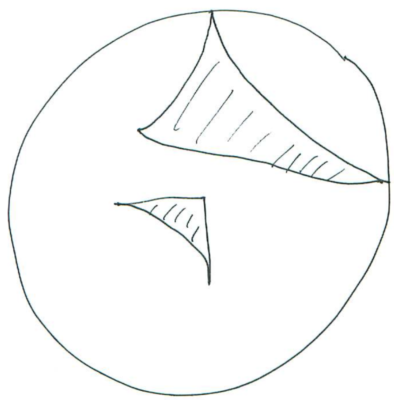
so that

$$\rho(w,0) = \log\frac{1+|w|}{1-|w|} = 2\tanh^{-1}|w|.$$

The map ϕ (27.7.15) maps the geodesics in $\mathbf{H}^2$ to geodesics in $\mathbf{D}^2$, and as a Möbius transformation, maps circles and lines to circles and lines, preserves angles, and maps the real axis to the unit circle; therefore the geodesics in $\mathbf{D}^2$ are diameters through the origin and semicircles orthogonal to the unit circle.



Accordingly, triangles from the upper half-plane map to triangles in the unit disc.



27.6.5. Via the map ϕ, the group $\mathrm{PSL}_2(\mathbb{R})$ acts on $\mathbf{D}^2$ as

$$\begin{aligned}\mathrm{PSL}_2(\mathbb{R})^\phi &= \phi\,\mathrm{PSL}_2(\mathbb{R})\phi^{-1} \\ &= \mathrm{PSU}(1,1) = \left\{\begin{pmatrix} u & v \\ \overline{v} & \overline{u}\end{pmatrix} \in \mathrm{M}_2(\mathbb{C}) : |u|^2 - |v|^2 = 1\right\}/\{\pm 1\}.\end{aligned}$$

Explicitly, an isometry of $\mathbf{D}^2$ is a map of the form

$$w \mapsto e^{i\theta}\left(\frac{w-a}{1-\overline{a}w}\right)$$

for $a \in \mathbb{C}$ with $|a| < 1$ and $\theta \in \mathbb{R}$. (A direct substitution can be used to give an alternate verification that these transformations are indeed isometries of $\mathbf{D}^2$ with the hyperbolic metric.)

The orientation reversing isometry $g(z) = -\overline{z}$ acts by $g^\phi(w) = \overline{w}$ with the choice $p = i$ (Exercise 27.12).

The induced area on $\mathbf{D}^2$ is given by

$$\mathrm{d}A = \frac{4\,\mathrm{d}x\,\mathrm{d}y}{(1-x^2-y^2)^2}$$

for $w = x + yi$.

Second, we present the Lorentz model.

Definition 27.6.6. The *Lorentz metric* on $\mathbb{R}^3$ is the indefinite metric

$$\mathrm{d}s^2 = -\,\mathrm{d}t^2 + \mathrm{d}x^2 + \mathrm{d}y^2.$$

27.6.7. The indefinite Lorentz metric is associated to the quadratic form

$$q(x, y, z) = -t^2 + x^2 + y^2$$

in the natural way. Lengths in this metric can be positive or nonpositive. However, on the *hyperboloid*

$$t^2 - x^2 - y^2 = 1,$$

the metric becomes positive definite: any nonzero tangent vector to the hyperboloid has positive length (Exercise 27.15). The hyperboloid can be thought of as the sphere of radius i about the origin with respect to q; taking an imaginary radius shows that hyperbolic geometry is dual in some sense to spherical geometry, where $\mathbf{S}^2 \subseteq \mathbb{R}^3$ has real radius 1.

Definition 27.6.8. The *Lorentz hyperboloid* is the set

$$\mathbf{L}^2 = \{(t, x, y) \in \mathbb{R}^3 : q(t, x, y) = -1, t > 0\}$$

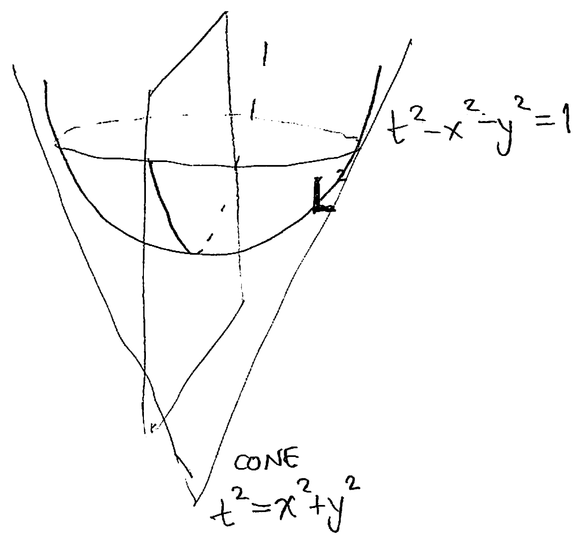
equipped with the Lorentz metric.

The Lorentz hyperboloid is the upper sheet of the (two-sheeted) hyperboloid; it is also called the *hyperboloid model* or the *Lorentz model* of planar hyperbolic geometry. (The choice of signs has to do with the physics of spacetime.)

The map

$$\begin{aligned} \mathbf{L}^2 &\to \mathbf{D}^2 \\ (t, x, y) &\mapsto (x + iy)/(t + 1) \end{aligned} \tag{27.6.9}$$

is bijective and identifies the metrics on $\mathbf{L}^2$ and $\mathbf{D}^2$ (Exercise 27.14). The map (27.6.9) maps geodesics in $\mathbf{D}^2$ to intersections of the hyperboloid with planes through the origin.



By pullback, we have $\mathrm{Isom}^+(\mathbf{L}^2) \simeq \mathrm{PSL}_2(\mathbb{R})$. However, other isometries are also apparent: a linear change of variables that preserves the quadratic form q also preserves the Lorentz metric. Let

$$\mathrm{O}(2,1) = \{g \in \mathrm{GL}_3(\mathbb{R}) : g^t m g = m\}, \quad \text{where } m = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Then

$$\mathrm{d}s^2 = v^t m v, \quad \text{where } v = (\mathrm{d}t, \mathrm{d}x, \mathrm{d}y)^t,$$

so if $g \in \mathrm{O}(2,1)$ then immediately $\mathrm{d}gs^2 = \mathrm{d}s^2$.

Next, if $g \in \mathrm{O}(2,1)$, then $g^t m g = m$ implies $\det(g) = \pm 1$. The elements of $\mathrm{O}(2,1)$ that map the hyperboloid to itself is the subgroup

$$\mathrm{SO}(2,1) = \{g \in \mathrm{O}(2,1) : \det(g) = 1\};$$

let $\mathrm{SO}^+(2,1) \leq \mathrm{SO}(2,1)$ be the further subgroup that maps the upper sheet of the hyperboloid to itself, the connected component of the identity.

Remark 27.6.10. We have proven that

$$\mathrm{SO}^+(2,1) \simeq \mathrm{PSL}_2(\mathbb{R}),$$

an exceptional isomorphism from Lie theory that also follows from the isomorphism of Lie algebras $\mathfrak{so}_{2,1} \simeq \mathfrak{sl}_2\mathbb{R}$.

27.7 Hyperbolic space

Definition 27.7.1. The *upper half-space* is the set

$$\mathbf{H}^3 = \{(z,t) = (x+iy,t) \in \mathbb{C} \times \mathbb{R} : t > 0\}.$$

Hyperbolic 3-space is the set $\mathbf{H}^3$ equipped with the metric induced by the *hyperbolic length element*

$$\mathrm{d}s^2 = \frac{\mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}t^2}{t^2}.$$

The space $\mathbf{H}^3$ is the unique three-dimensional (connected and) simply connected Riemannian manifold with constant sectional curvature -1. The *volume element* corresponding to the hyperbolic length element is accordingly

$$\mathrm{d}V = \frac{\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}t}{t^3}.$$

The *sphere at infinity* is the set

$$\partial\mathbf{H}^3 = \mathbb{P}^1(\mathbb{C}) = \mathbb{C} \cup \{\infty\}$$

(analogous to the circle at infinity for $\mathbf{H}^2$), with the image of $\mathbb{C}$ corresponding to the locus of points with $t = 0$. We then define the *completed upper half-space* to be

$$\mathbf{H}^{3*} = \mathbf{H}^3 \cup \partial\mathbf{H}^3.$$

The topology on $\mathbf{H}^{3*}$ is defined by taking a fundamental system of neighborhoods of the point at ∞ to be sets of the form

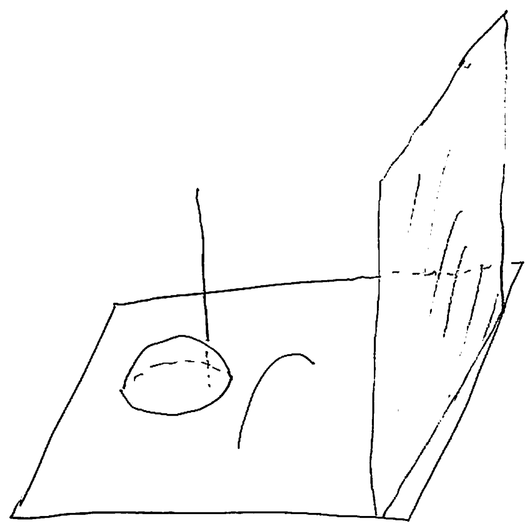
$$\{(z,t) \in \mathbf{H}^3 : t > M\} \cup \{\infty\}$$

for $M > 0$.

27.7.2. The metric space $\mathbf{H}^3$ is complete, and the topology on $\mathbf{H}^3$ is the same as the topology induced by the Euclidean metric.

The bi-infinite geodesics in $\mathbf{H}^3$ are the Euclidean hemicircles orthogonal to $\mathbb{C}$ and vertical half-lines: indeed, any two points lie in a vertical plane in which the restriction of the metric is equivalent to the hyperbolic metric, so this statement can be deduced from the case of the hyperbolic plane. (Alternatively, by applying an element of $\mathrm{PSL}_2(\mathbb{C})$ it is enough to show that the vertical axis $Z = \{(0,0,t) : t > 0\}$ is a geodesic, and arguing as in (27.4.4) we obtain the result.) Accordingly, $\mathbf{H}^3$ is a geodesic space, and there is a unique geodesic between any two points.

The geodesic planes in $\mathbf{H}^3$ are the Euclidean hemispheres orthogonal to $\mathbb{C}$ and the vertical half-planes.



27.7.3. Analogous to the case of $\mathbf{H}^2$, we find a group of isometries of $\mathbf{H}^3$ defined by linear fractional transformations by $\mathrm{PSL}_2(\mathbb{C})$. We first describe these geometrically, via the *Poincaré extension*. An element $g \in \mathrm{SL}_2(\mathbb{C})$ as a Möbius transformation, induces a biholomorphic map of the Riemann sphere $\mathbb{P}^1(\mathbb{C}) = \mathbb{C} \cup \{\infty\}$. This map can be represented as a composition of an even number (at most four) inversions in circles in $\mathbb{P}^1(\mathbb{C})$, or circles and lines in $\mathbb{C}$ (Exercise 27.17). We have identified $\mathbb{P}^1(\mathbb{C}) = \partial\mathbf{H}^3$ as the boundary, and for each circle in $\mathbb{P}^1(\mathbb{C})$ there is a unique hemisphere in $\mathbf{H}^3$ which intersects $\partial\mathbf{H}^3$ in this circle; if this circle is a line, then we take a vertical half-plane. We then lift the action of $g \in \mathrm{PSL}_2(\mathbb{C})$ one inversion at a time, with respect to the appropriate hemisphere or half-plane. It turns out that the action of this product does not depend on the choice of the circles.

27.7.4. The geometric action (27.7.3) of $\mathrm{PSL}_2(\mathbb{C})$ can be described by formulas. The action on the boundary $\mathbb{P}^1(\mathbb{C})$ is given by

$$g(z) = \frac{az+b}{cz+d}.$$

We then identify

$$\begin{aligned} \mathbf{H}^3 &\hookrightarrow \mathbb{H} = \mathbb{C} + \mathbb{C}j \\ (z,t) &\mapsto u = z + tj \end{aligned}$$

where we recall that $jz = \overline{z}j$ for $z \in \mathbb{C} = \mathbb{R} + \mathbb{R}i \subseteq \mathbb{H}$; it is then natural define the map

$$\begin{aligned} \mathrm{SL}_2(\mathbb{C}) \times \mathbf{H}^3 &\to \mathbb{H}^\times \\ (g,u) &\mapsto gu = (au+b)(cu+d)^{-1} \end{aligned} \tag{27.7.5}$$

for $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{C})$.

Lemma 27.7.6. *The map* (27.7.5) *defines a faithful, transitive action of* $\mathrm{PSL}_2(\mathbb{C})$ *on* $\mathbf{H}^3$.

Proof. The fact that (27.7.5) defines an action can be checked by a calculation: If $g(u) = g(z+tj) = z' + t'j = u'$, then in coordinates we have (Exercise 27.18)

$$\begin{aligned} z' &= \frac{(az+b)\overline{(cz+d)} + a\bar{c}t^2}{\|cw+d\|^2} \\ t' &= \frac{t}{|cz+d|^2 + |c|^2 t^2} = \frac{t}{\|cw+d\|^2} \end{aligned} \tag{27.7.7}$$

where

$$\|cw+d\|^2 = \mathrm{nrd}(cw+d)^2 = |cz+d|^2 + |c|^2 t^2.$$

For transitivity, we show that $\mathbf{H}^3$ is the orbit of j. Indeed, if $u = z + tj \in \mathbf{H}^3$ then we first apply a translation $\begin{pmatrix} 1 & -z \\ 0 & 1 \end{pmatrix}$ to get $w = tj$ and then reduce to the case of the hyperbolic plane. $\square$

Next, we show the action of $\mathrm{PSL}_2(\mathbb{C})$ on $\mathbf{H}^3$ is by isometries. We verify this on a convenient set of generators as follows.

Lemma 27.7.8. *The group* $\mathrm{SL}_2(\mathbb{C})$ *is generated by the elements*

$$\left\{ \begin{pmatrix} a & 0 \\ 0 & 1/a \end{pmatrix} : a \in \mathbb{C}^\times \right\} \cup \left\{ \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix} : b \in \mathbb{C} \right\} \cup \left\{ \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \right\}.$$

Proof. Let $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{C})$. If $c \neq 0$, then

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 1 & a/c \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} c & 0 \\ 0 & 1/c \end{pmatrix} \begin{pmatrix} 1 & d/c \\ 0 & 1 \end{pmatrix}$$

and if $c = 0$ we have

$$\begin{pmatrix} a & b \\ 0 & 1/a \end{pmatrix} = \begin{pmatrix} a & 0 \\ 0 & 1/a \end{pmatrix} \begin{pmatrix} 1 & b/a \\ 0 & 1 \end{pmatrix}.$$ $\square$

Remark 27.7.9. In fact, the generators $\begin{pmatrix} a & 0 \\ 0 & 1/a \end{pmatrix}$ are redundant, but we will not use this fact here.

Proposition 27.7.10. *We have*

$$\mathrm{Isom}^+(\mathbf{H}^3) \simeq \mathrm{PSL}_2(\mathbb{C})$$

and

$$\mathrm{Isom}(\mathbf{H}^3) \simeq \mathrm{PSL}_2(\mathbb{C}) \rtimes \mathbb{Z}/2\mathbb{Z}$$

where the nontrivial element of $\mathbb{Z}/2\mathbb{Z}$ *acts by complex conjugation on* $\mathrm{PSL}_2(\mathbb{C})$ *and* $(z,t) \mapsto (\overline{z}, t)$ *on* $\mathbf{H}^3$.

Proof. To show that $\mathrm{PSL}_2(\mathbb{C}) \hookrightarrow \mathrm{Isom}(\mathbf{H}^3)$, by Lemma 27.7.8, it suffices to consider for $u = z + tj$ the elements

$$\begin{pmatrix} a & 0 \\ 0 & 1/a \end{pmatrix}(u) = a^2 z + |a|^2 tj,$$
$$\begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix}(u) = (z+b) + tj,$$
$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}(u) = \frac{1}{\|u\|}(-\overline{z} + tj)$$

where again $\|u\| = |z|^2 + t^2$. Verification that $\mathrm{d}g(s) = \mathrm{d}s$ and that orientation is preserved in the first two cases is immediate from the definition of the metric, and the third case follows inversion in the unit sphere together with complex conjugation (Exercise 27.19) and can checked directly.

Now let $\phi \in \mathrm{Isom}(\mathbf{H}^3)$. Then ϕ maps geodesic hemispheres and vertical half-planes to the same, and consequently by reversing the Poincaré extension, it maps circles and lines in $\mathbb{C}$ to the same. The only continuous functions that have this property are complex conjugation and the Möbius transformations, so the final result follows. ☐

Alternatively, another way to proceed is via the description in Paragraph 27.7.3: it suffices to show that inversion in a hemisphere preserves the hyperbolic metric, and to show this, together with the first two types of generators in Lemma 27.7.8, it suffices to check that inversion in the unit hemisphere is a hyperbolic isometry, essentially the same computation as in checking this for $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$.

Lemma 27.7.11. *The stabilizer of $j \in \mathbf{H}^3$ in $\mathrm{SL}_2(\mathbb{C})$ is equal to*

$$\mathrm{SU}(2) = \left\{ \begin{pmatrix} a & b \\ -\overline{b} & \overline{a} \end{pmatrix} \in \mathrm{M}_2(\mathbb{C}) : |a|^2 + |b|^2 = 1 \right\}.$$

Proof. From (27.7.7), we see that $gi = (ai+b)(ci+d)^{-1}$ if and only if $|c|^2 + |d|^2 = 1$ and $a\overline{c} + b\overline{d} = 0$. Plugging the first equation into the second, and using $ad - bc = 1$ gives $a = \overline{d}$ and then $b = -\overline{c}$. ☐

27.7.12. The group $\mathrm{PSL}_2(\mathbb{C})$ acts transitively on geodesics and consequently on pairs of points at a fixed distance: by the transitive action of $\mathrm{PSL}_2(\mathbb{C})$ on $\mathbf{H}^3$, any point can be mapped to j; and applying an element of $\mathrm{Stab}_{\mathrm{SL}_2(\mathbb{C})}\, j = \mathrm{SU}(2)$, any other point u can be brought to tj with $t \geq 1$, with $\log t = \rho(j, u)$ by the distance in the hyperbolic plane. It follows that

$$\cosh \rho(u, u') = \frac{|z - z'|^2 + t^2 + t'^2}{2tt'} \tag{27.7.13}$$

by verifying (27.7.13) in the special case where $u = j$ and $u' = tj$ with $t \geq 1$, and then using the preceding transitive action and the fact that the the right-hand side of (27.7.13) is invariant under the action of $\mathrm{SL}_2(\mathbb{C})$, verified using the generators in Lemma 27.7.8.

We have a similar classification of isometries as follows. Let $g \in \mathrm{PSL}_2(\mathbb{C})$. If $\pm\mathrm{Tr}(g) \in (-2,2)$, then g is *elliptic*: it has two distinct fixed points in $\partial\mathbf{H}^3$ and fixes every point in the geodesic between them, called its *axis*. If $\pm\mathrm{Tr}(g) \in \mathbb{C} \setminus [-2,2]$, then g is *loxodromic*: it has two fixed points in $\partial\mathbf{H}^3$ and stabilizes the geodesic axis between them. Finally, if $\pm\mathrm{Tr}(g) = \pm 2$, then g is *parabolic*: it has a unique fixed point in $\partial\mathbf{H}^3$ and no fixed point in $\mathbf{H}^3$. [[Iwasawa decomposition here?]]

Definition 27.7.14. The *hyperbolic unit ball* is the (open) unit disc

$$\mathbf{D}^3 = \{w = (w_1, w_2, w_3) \in \mathbb{R}^3 : \|w\|^2 < 1\}$$

equipped with the *hyperbolic metric*

$$\mathrm{d}s = \frac{2\|\mathrm{d}w\|}{1 - \|w\|^2}$$

and volume

$$\mathrm{d}V = 8\frac{\mathrm{d}w_1\,\mathrm{d}w_2\,\mathrm{d}w_3}{(1 - \|w\|)^3}.$$

The *sphere at infinity* is the boundary

$$\partial\mathbf{D}^3 = \{w \in \mathbb{R}^3 : \|w\| = 1\}.$$

27.7.15. The maps

$$\begin{aligned} \phi : \mathbf{H}^3 &\xrightarrow{\sim} \mathbf{D}^3 \\ u &\mapsto w = (u - j)(1 - ju)^{-1} \end{aligned} \qquad \begin{aligned} \phi^{-1} : \mathbf{D}^3 &\xrightarrow{\sim} \mathbf{H}^3 \\ w &\mapsto u = (w + j)(1 + jw)^{-1} \end{aligned}$$

define a conformal equivalence between $\mathbf{H}^3$ and $\mathbf{D}^3$ with $j \mapsto \phi(j) = 0$. The hyperbolic metric on $\mathbf{D}^2$ is the pushforward of (induced from) the hyperbolic metric on $\mathbf{H}^3$ via the identification (27.7.15). We find that

$$\cosh\rho(w, w') = 1 + 2\frac{\|w - w'\|^2}{(1 - \|w\|^2)(1 - \|w'\|^2)}. \tag{27.7.16}$$

In the unit ball model, the (bi-infinite) geodesics are intersections of $\mathbf{D}^3$ of Euclidean circles and straight lines orthogonal to the sphere at infinity, and similarly geodesic planes are intersections of $\mathbf{D}^3$ with Euclidean spheres and Euclidean planes orthogonal to the sphere at infinity.

27.7.17. The isometries of $\mathbf{D}^3$ are obtained by pushforward from $\mathbf{H}^3$. Explicitly, we first identify

$$\begin{aligned} \mathbf{D}^3 &\hookrightarrow \mathbb{H} \\ w &\mapsto w_1 + w_2 i + w_3 j. \end{aligned}$$

We then define the involution

$$\begin{aligned} {}^* : \mathbb{H} &\to \mathbb{H} \\ \alpha = t + xi + yj + zk &\mapsto k\overline{\alpha}k^{-1} = t + xi + yj - zk \end{aligned}$$

and the group

$$\mathrm{SU}_2(\mathbb{H}, ^*) = \left\{ \begin{pmatrix} u & v \\ v^* & u^* \end{pmatrix} : \mathrm{nrd}(u) - \mathrm{nrd}(v) = 1 \right\}.$$

We find that

$$\mathrm{SU}_2(\mathbb{H}, ^*) \simeq \phi\,\mathrm{SL}_2(\mathbb{C})\phi^{-1}$$

with ϕ as in Paragraph 27.7.15. The group $\mathrm{SU}_2(\mathbb{H}, ^*)$ acts on D^3 by

$$gw = (uw + v)(v^*w + u^*)^{-1}.$$

27.7.18. Finally, we have a *Lorentz model*

$$\mathsf{L}^3 = \{(t, x) \in \mathbb{R}^4 : -t^2 + x_1^2 + x_2^2 + x_3^2 = -1, t > 0\}$$

with

$$\mathrm{d}s^2 = -\,\mathrm{d}t^2 + \mathrm{d}x_1^2 + \mathrm{d}x_2^2 + \mathrm{d}x_3^2$$

and orientation-preserving isometries given by the subgroup $\mathrm{SO}^+(3,1) \leq \mathrm{SO}^+(3,1)$ of elements mapping L^3 to itself. The relationship between the Lorentz model and the upper half-space model relies on the exceptional isomorphism in Lie theory given by $\mathrm{SO}(3,1) \simeq \mathrm{SL}_2(\mathbb{C})$.

27.8 Some Riemannian geometry

The hyperbolic metric (27.2.3) is induced from a Riemannian metric as follows.

A *Riemannian metric* $\mathrm{d}s^2$ on an open set $U \subseteq \mathbb{R}^n$ is function that assigns to each point $p \in U$ a (symmetric, positive definite) inner product on the tangent space T_p at $p \in U$, varying differentiably. Such an inner product defines the length of a tangent vector, the angle between two tangent vectors, and the length element $\mathrm{d}s = \sqrt{\mathrm{d}s^2}$. In coordinates, we write

$$\mathrm{d}s^2 = \sum_{i,j} \eta_{ij}\,\mathrm{d}x_i\,\mathrm{d}x_j$$

from standard coordinates x_i on $\mathbb{R}^n$, and the matrix (g_{ij}) is symmetric, positive definite, and differentiable in x. The metric determines a volume formula as

$$\mathrm{d}V = \sqrt{\det \eta}\,\mathrm{d}x_1 \cdots \mathrm{d}x_n.$$

If $\phi : \mathbb{R}^k \to \mathbb{R}^n$ is differentiable, the pullback metric $\phi^*(\mathrm{d}s^2)$ is defined by

$$\phi^*(\mathrm{d}s^2)(v, w) = \mathrm{d}s^2(Df(v), Df(w))$$

where $v, w \in T_p$ and Df is the derivative map.

The tangent space to $\mathcal{H}$ at a point $z \in \mathcal{H}$ is $\mathrm{T}_z\,\mathcal{H} \simeq \mathbb{C}$. Therefore, the tangent bundle

$$\mathrm{T}(\mathcal{H}) = \{(z, v) : z \in \mathcal{H}, v \in \mathrm{T}_z\,\mathcal{H}\}$$

is trivial (*parallelizable*), with $\mathrm{T}(\mathcal{H}) \simeq \mathcal{H} \times \mathbb{C}$. The Riemannian metric on $\mathcal{H}$ is then defined by the metric on $\mathrm{T}_z\,\mathcal{H}$ given by

$$\langle v, w\rangle = \frac{\operatorname{Re} v \operatorname{Re} w + \operatorname{Im} v \operatorname{Im} w}{(\operatorname{Im} z)^2}$$

for $v, w \in \mathrm{T}_z\,\mathcal{H}$—we just rescaled the usual inner product on $\mathbb{C}$. In particular, $\|v\|_z = |v|/(\operatorname{Im} z)$. The angle between two geodesics at an intersection point $z \in \mathrm{T}_z\,\mathcal{H}$ is then defined to be the angle between their tangent vectors in $\mathrm{T}_z\,\mathcal{H}$; this notion of an angle coincides with the Euclidean angle measure.

The action of $\mathrm{PSL}_2(\mathbb{R})$ on $\mathcal{H}$ extends to an action on $T\mathcal{H}$ in the expected way:

$$g(z, v) = (gz, g'(z)v) = \left(\frac{az + b}{cz + d}, \frac{1}{(cz + d)^2}v\right).$$

Since isometries of $\mathcal{H}$ are differentiable, they act on the tangent bundle by differentials preserving the norm and angle, and therefore $\operatorname{Isom}(\mathcal{H})$ acts conformally or anti-conformally on $\mathcal{H}$.

If we restrict to the *unit tangent bundle*

$$\mathrm{UT}(\mathcal{H}) = \{(z, v) \in \mathrm{T}(\mathcal{H}) : \|v\|_z^2 = 1\}$$

then we obtain a bijection

$$\begin{aligned} \mathrm{PSL}_2(\mathbb{R}) &\xrightarrow{\sim} \mathrm{UT}(\mathcal{H}) \\ g &\mapsto (gi, g'(i)i) \end{aligned}$$

(Exercise 27.16).

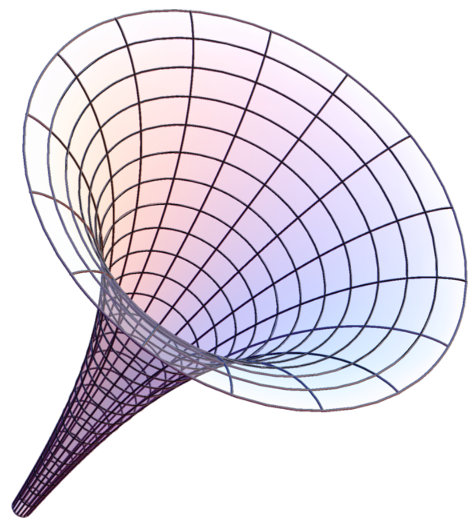
Similar statements hold for $\mathrm{PSL}_2(\mathbb{C})$ in place of $\mathrm{PSL}_2(\mathbb{R})$.

27.9 Extensions and further reading

The underpinnings of what became hyperbolic geometry can be found in work by Euler and Gauss in their study of curved surfaces (the differential geometry of surfaces). It was then Lobachevsky and Bolyai who suggested that curved surfaces of constant negative curvature could be used in non-Euclidean geometry, and finally Riemann who generalized this to what are now called Riemmanian manifolds. Hyperbolic geometry, and in particular the hyperbolic plane, remains an important prototype for understanding negatively-curved spaces in general. Milnor [Mil82] gives a comprehensive early history of hyperbolic geometry; see also the survey by Cannon–Floyd–Kenyon–Parry.

Anderson [And2005], Katok [Kat92, Chapter 1] and Beardon [Bea95, Chapter 7] provide further references for hyperbolic plane geometry. There are a wealth of geometric results and formulas from Euclidean geometry that one can try to reformulate in the world of hyperbolic plane geometry, and the interested reader is encouraged to pursue these further.

One way to visualize plane hyperbolic geometry by the *pseudosphere*, the surface of revolution generated by a tractrix: it is a surface with constant negative curvature and so locally models the hyperbolic plane.



Elstrodt–Grunewald–Mennicke [EGM98] Marden [Mar2007] are references for the geometry of hyperbolic space.

The natural generalization of geometry in the classical sense is performed on a Riemannian manifold X that is *homogeneous*, so the isometry group $\mathrm{Isom}(X)$ acts transitively on X and every point "looks the same", as well as *isotropic*, so $\mathrm{Isom}(X)$ acts transitively on frames (a basis of orthonormal tangent vectors) at a point and the geometry "looks the same" in any direction at a point. Taken together, these natural conditions are quite strong, and there are only three essentially distinct simply connected homogeneous and isotropic geometries in any dimension, corresponding to constant sectional curvatures zero, positive, or negative: these are Euclidean, spherical, and hyperbolic geometry, respectively. Put this way, the hyperbolic plane is the unique complete, simply connected Riemann surface with constant sectional curvature -1. For more on geometries in this sense, we encourage the reader to consult Thurston [Thu97].

27.9.1. Theorem 27.5.5 is called the Gauss–Bonnet formula because it is closely related to the more general formula relating curvature to Euler characteristic. The simplest kind of formula of this kind is

$$\int_X K \,\mathrm{d}A = 2\pi\chi(X) \tag{27.9.2}$$

for a Riemann surface X. The expression (27.9.2) is quite remarkable: it says that the total curvature of X is determined by its topology; if you flatten out a surface in one place, the curvature is forced to rise somewhere else. If instead one has a surface X with geodesic boundary, then the formula (27.9.2) becomes

$$\int_X K \,\mathrm{d}a + \sum_i (\pi - \theta_i) = 2\pi\chi(X)$$

where θ_i are the angles at the vertices. For a triangle X with constant curvature -1 and angles α, β, γ, we have $\int_X K\,\mathrm{d}A = -\mu(X)$ and $\chi(X) = V - E + F = 1$ (like any polygon), so we find

$$-\mu(X) + 3\pi - (\alpha + \beta + \gamma) = 2\pi$$

and we recover Theorem 27.5.5.

27.9.3. More generally, one defines *hyperbolic upper half-space*

$$\mathbf{H}^n = \{x \in \mathbb{R}^n : x_n > 0\} \text{ with } \mathrm{d}s^2 = \frac{\mathrm{d}x_1^2 + \cdots + \mathrm{d}x_n^2}{x_n^2}.$$

The space $\mathbf{H}^n$ is a geodesic space and a model for *hyperbolic n-space*. The geodesics in $\mathbf{H}^n$ are orthocircles, and via a conformal map. The upper half-space maps isometrically to the (open) unit ball model

$$\mathbf{D}^n = \{x \in \mathbb{R}^n : |x| < 1\} \text{ with } \mathrm{d}s^2 = 4\frac{\mathrm{d}x_1^2 + \cdots + \mathrm{d}x_n^2}{(1 - x_1^2 - \cdots - x_n^2)^2}$$

and the hyperboloid model

$$\mathbf{L}^n = \{(t, x) \in \mathbb{R}^{n+1} : -t^2 + x_1^2 + \cdots + x_n^2 = -1,\ t > 0\}$$

with

$$\mathrm{d}s^2 = -\,\mathrm{d}t^2 + \mathrm{d}x_1^2 + \cdots + \mathrm{d}x_n^2.$$

These models (and more) are introduced and compared in Cannon–Floyd–Kenyon–Parry [CFKP97], and treated in detail in the works by Benedetti–Petronio [BP92] and Ratcliffe [Rat2006].

27.9.4. If a metric space X is complete and locally compact, then by the Hopf–Rinow theorem, any two points in X can be connected by a minimizing geodesic and all bounded closed sets in X are compact.

Exercises

27.1. Show that the hyperbolic metric is equivalent to the Euclidean metric in two ways.

a) Show directly that open balls nest: for all $z \in \mathcal{H}$ and all $\epsilon > 0$, there exist $\eta_1, \eta_2 > 0$ such that

$$\rho(z, w) < \eta_1 \Rightarrow |z - w| < \epsilon \Rightarrow \rho(z, w) < \eta_2$$

for all $w \in \mathcal{H}$.

b) Show that the collection of Euclidean balls coincides with the collection of hyperbolic balls. *[Hint: applying an isometry, reduce to the case of balls around i and check this directly; it is perhaps even clearer moving to the unit disc model.]*

27.2. Check that in $\mathbb{R}^n$, the metric specified in (27.1.6)

$$\ell(v) = \int_v \sqrt{x_1'(t)^2 + \cdots + x_n'(t)^2}\ \mathrm{d}t$$

has lines as geodesics.

27.3. From differential geometry, the curvature of a Riemann surface with metric

$$ds = \sqrt{f(x,y)\,\mathrm{d}x^2 + g(x,y)\,\mathrm{d}y^2}$$

is given by the formula

$$-\frac{1}{\sqrt{fg}}\left(\frac{\partial}{\partial x}\left(\frac{1}{\sqrt{f}}\frac{\partial\sqrt{g}}{\partial x}\right) + \frac{\partial}{\partial y}\left(\frac{1}{\sqrt{g}}\frac{\partial\sqrt{f}}{\partial y}\right)\right)$$

for suitably nice functions f, g. Using this formula, verify that the curvature of $\mathbf{H}^2$ and $\mathbf{D}^2$ is -1.

27.4. Consider $\mathbb{C}$ with the standard metric. Let

$$\mathrm{Isom}^h(\mathbb{C}) = \{g \in \mathrm{Isom}(\mathbb{C}) : g \text{ is holomorphic}\} \leq \mathrm{Isom}(\mathbb{C}).$$

Show that

$$\mathrm{Isom}^h(\mathbb{C}) \simeq \left\{\begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix} \in \mathrm{GL}_2(\mathbb{C}) : |a| = 1\right\}$$

and exhibit an isometry of $\mathbb{C}$ that is not holomorphic.

27.5. Show that for any two points $z, w \in \mathcal{H}$, there exists $g \in \mathrm{PSL}_2(\mathbb{R})$ such that $gz, gz' \in \mathbb{R}_{>0}i$ are pure imaginary.

27.6. Show that the image of $\mathbb{R}_{>0}i$ under an element of $\mathrm{PSL}_2(\mathbb{R})$ is either a semicircle orthogonal to $\mathbb{R}$ or a vertical line.

27.7. We consider the action of $\mathrm{PSL}_2(\mathbb{R})$ on geodesics in $\mathcal{H}$.

a) Show that $\mathrm{PSL}_2(\mathbb{R})$ acts transitively on the set of bi-infinite geodesics in $\mathcal{H}$.

b) Show that given any two points $z, w \in \mathcal{H}$, there exists $g \in \mathrm{PSL}_2(\mathbb{R})$ such that $gz = w$ and such that g that maps the bi-infinite geodesic through z and w to itself.

c) Show that the isometry of $\mathcal{H}$ that maps a geodesic to itself and fixes two points on this geodesic is the identity.

27.8. Show that the expression

$$\log \frac{|z - \overline{w}| + |z - w|}{|z - \overline{w}| - |z - w|}$$

with $z, w \in \mathbf{H}^2$ is invariant under $g \in \mathrm{PSL}_2(\mathbb{R})$.

27.9. Show that

$$\cosh \log \frac{|z-\overline{w}|+|z-w|}{|z-\overline{w}|-|z-w|} = 1 + \frac{|z-w|^2}{2\operatorname{Im}(z)\operatorname{Im}(w)}$$

for all $z, w \in \mathbf{H}^2$.

27.10. Show that the Gauss–Bonnet formula (Theorem 27.5.5) holds when T has two or three vertices on $\partial\mathcal{H}$.

27.11. Verify that the hyperbolic metric on $\mathbf{D}^2$ is induced from the hyperbolic metric on $\mathbf{H}^2$ from the identification (27.6.4), as follows.

a) Show that

$$\frac{2|\phi'(z)|}{1-|\phi(z)|^2} = \frac{1}{\operatorname{Im} z}.$$

b) Let $w = f(z)$, and conclude that

$$\frac{2|\mathrm{d}w|}{1-|w|^2} = \frac{|\mathrm{d}z|}{\operatorname{Im} z}.$$

27.12. Show that the orientation-reversing isometry $g(z) = -\overline{z}$ induces the map

$$g^{\phi}(w) = \overline{w}$$

via the conformal transformation $\phi : \mathbf{H}^2 \to \mathbf{D}^2$ in (27.6.4).

27.13. Show that the Iwasawa decomposition (Proposition 27.3.2) can be given explicitly as

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 1 & (ac+bd)/r \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1/r & 0 \\ 0 & r \end{pmatrix} \begin{pmatrix} s & -t \\ t & s \end{pmatrix} \in NAK = \mathrm{SL}_2(\mathbb{R})$$

where $r = \sqrt{c^2+d^2}$, $s = d/r$, $t = c/r$.

27.14. Show that the map

$$\begin{aligned} \mathbf{L}^2 &\to \mathbf{D}^2 \\ (t,x,y) &\mapsto (x+iy)/t \end{aligned}$$

identifies the metrics on $\mathbf{L}^2$ and $\mathbf{D}^2$.

27.15. Show that the Lorentz metric restricted to the hyperboloid is an honest (Riemannian) metric. *[Hint: Show that a tangent vector v at a point p satisfies $b(p,v) = 0$, where b is the quadratic form associated to q; show that the orthogonal complement to p has signature $+2$.]*

27.16. Show that there is a bijection $\mathrm{PSL}_2(\mathbb{R}) \xrightarrow{\sim} \mathrm{UT}(\mathcal{H})$ defined by the action of g on a fixed base point in $\mathrm{UT}(\mathcal{H})$ (such as (i,i)). *[Hint: Observe that elliptic elements rotate the tangent vector.]*

27.17. Verify that every element of $\mathrm{SL}_2(\mathbb{C})$ can be written as a composition of (at most four) inversions in circles in $\mathbb{P}^1(\mathbb{C})$, or circles and lines in $\mathbb{C}$.

27.18. Verify (27.7.7) for the action of $\mathrm{SL}_2(\mathbb{C})$ on $\mathbf{H}^3 \subseteq \mathbb{H}$.

27.19. Let $g = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \in \mathrm{SL}_2(\mathbb{C})$. Show that g acts on $\mathbf{H}^3$ by

$$gu = \frac{1}{\|u\|}(-\overline{z}, t)$$

where $\|u\| = |z|^2 + t^2$, and that g is a hyperbolic isometry.