The goal of the second set of topics in the course is to introduce students to hyperbolic geometry and its connections to dynamical systems and number theory. We build on abstract notions of topological spaces and groups that come together in this beautiful and very concrete subject. We will mostly deal with isometries of the hyperbolic plane, with projects covering connections to number theory and dynamical systems.

We begin by recalling several ways of looking at $\mathbb{R}^2$ (different models) and two notions that we generalise.

\begin{align*}
\mathbb{R}^2 &= \mathbb{R} \times \mathbb{R} \text{ (Cartesian coordinates)} \quad (1) \\
&= \mathbb{C} = \mathbb{R} \times i\mathbb{R} \text{ (complex numbers)} \quad (2) \\
&= \{(r, \alpha) : r > 0, \alpha \in [0, 2\pi) \text{ or } r = 0, \alpha \text{ arbitrary}\} \text{ (polar coordinates)} \quad (3) \\
&= S^2 \setminus \{\infty\} \text{ (two-sphere with one point removed).} \quad (4)
\end{align*}

The two notions we need are arc length and area. The arc length of a piecewise $C^1$ curve $\gamma = (x, y) : [0, 1] \to \mathbb{R}^2$ is given by

$$
\int_0^1 \|\gamma'(t)\|dt,
$$

(5)

where $\|\cdot\|$ denotes Euclidean norm. The term piecewise $C^1$ means that $\gamma$ is continuous (i.e., $C^0$) and its derivative is continuous except at finitely many points. Writing everything out in Cartesian coordinates, we have

$$
\int_0^1 \|\gamma'(t)\|dt = \int_0^1 \sqrt{(\frac{dx}{dt})^2 + (\frac{dy}{dt})^2} dt = \int \sqrt{(dx)^2 + (dy)^2} = \int ds.
$$

(6)

d$s$ is called the arc length element; in $\mathbb{R}^2$ it satisfies $ds^2 = dx^2 + dy^2$, and parentheses are customarily omitted. For any region $S \subset \mathbb{R}^2$ (with piecewise $C^1$ boundary), the area is computed via

$$
\int_S dx \, dy = \int_S dA.
$$

(7)

So the area element is $dA = dx \, dy$.

The object of our study is the hyperbolic plane, $\mathbb{H}^2$, which we define by $\{(x, y) : x \in \mathbb{R}, y > 0\}$ with the following arc length and area elements:

$$
ds^2 = \frac{dx^2 + dy^2}{y^2} \quad \quad \quad \quad \quad dA = \frac{dx \, dy}{y^2}.
$$

(8)
Equivalently, we can write \( \{ x + iy : x \in \mathbb{R}, y > 0 \} = \mathbb{H}^2 \) or \( \{ z \in \mathbb{C} : \text{Im} \, z > 0 \} \). These are of course equivalent, and letters \( x, y, \) and \( z \) should be assumed be in one of these realisations.

Example 1. Let \( \gamma(t) = (t, T) \) for a fixed \( T > 0 \) and \( t \in [0, 1] \). The length of \( \gamma \) is

\[
\ell(\gamma) = \int_0^1 \frac{\sqrt{t^2 + 0^2}}{T} \, dt = \frac{1}{T},
\]

so that curve is short for large \( T \) and long for \( T \) close to 0.

Example 2. Let \( S = [0, 1] \times [1, \infty) = \{ 0 \leq x \leq 1, y \geq 1 \} \). The area of \( S \) is

\[
A(S) = \int_1^\infty \int_0^1 \frac{dy}{y^2} = 1.
\]

It helps to sketch the region to establish the limits in general.

Example 3. Let \( \gamma(t) = (1, t) \) for \( t \in (0, 1) \). Then

\[
\ell(\gamma) = \int_0^1 \frac{\sqrt{0^2 + 1^2}}{t} \, dt = \int_0^1 \frac{dt}{t} = \infty.
\]

Note that the point \( (1, 0) \) is not in \( \mathbb{H}^2 \), so this curve goes to the boundary of \( \mathbb{H}^2 \) (defined rigorously later).

Example 4. Let \( S = \{ x^2 + y^2 < 1 \} \). Assuming (as we should) that \( y > 0 \), the area of \( S \) is

\[
A(S) = \int_{x=-1}^1 \int_{y=0}^{\sqrt{1-x^2}} \frac{dy}{y^2} \, dx = \int_{-1}^1 \infty \, dx = \infty.
\]

Performing the integration in the opposite order gives the same answer:

\[
A(S) = \int_{y=0}^{\sqrt{1-y^2}} \int_{x=-\sqrt{1-y^2}}^{1} \frac{dx}{y^2} \, dy = \int_0^{\sqrt{1-y^2}} 2\frac{1-y^2}{y^2} \, dy = \infty.
\]

Again, this region touches the line \( \{ y = 0 \} \) (which is not in the space) “a lot,” so it is reasonable for its area to be infinite.

Now we define a metric on the space \( \mathbb{H}^2 \). This is a standard construction and should work for a general arc length differential \( ds \). Set

\[
d(p, q) = \inf_{\gamma : [0,1] \to \mathbb{H}^2} \ell(\gamma),
\]

where

\[
\gamma(0) = p, \gamma(1) = q, \quad \gamma \text{ piecewise } C^1
\]

In the space we work with the above infimum is achieved, as we shall see, by a unique curve among curves of constant speed.
Theorem 1. $d$ is a metric.

Proof. We need to prove that $d$ is reflexive, symmetric, and transitive (part in class, part for homework). 

A curve that realises the infimum in the definition of $d$ is call a geodesic.

Example. Is there a geodesic that joins $i$ to $ie$? Well, for any curve $\gamma(t) = (x(t), y(t))$ we can write

$$
\int_0^1 \frac{\sqrt{(x'(t))^2 + (y'(t))^2}}{y(t)} dt \geq \int_0^1 \frac{|y'(t)|}{y(t)} dt \geq \int_0^1 \frac{y'(t)}{y(t)} dt = \int_0^1 \frac{dy}{y} = 1.
$$

So, the distance is at least 1. How do we achieve this distance? The first inequality is an equality precisely when $x'(t) = 0$, so we must assume $x(t)$ is a constant. Since $x(0) = 0$, $x(t) = 0$ for all $t$. The second inequality is an equality only if $|y'(t)| = y'(t)$; i.e., $y'(t) \geq 0$ for all $t$. Hence any curve that starts at $i$ and goes straight up (possibly stopping along the way but never backtracking) is a geodesic. The image $\gamma([0, 1]) = \{x = 0, 1 \leq y \leq e\} \subset \mathbb{H}^2$ is the same for any such curve. Note that since all assumptions we made about the minimising curve were forced (check this!), there is a unique geodesic of constant speed, namely $\gamma(t) = (0, e^t)$, with speed equal to

$$
\frac{\sqrt{(x'(t))^2 + (y'(t))^2}}{y(t)} = 1
$$

for every $t$.

One could perform a similar (but much more tedious) calculation for any two points and show that there is a unique geodesic joining them. We do this more elegantly later by studying the group of isometries of $\mathbb{H}^2$. Some definitions first.

An action of a group $G$ on a set $X$ (denoted $G \curvearrowright X$) is a map

$$
G \times X \to X
$$

$$(g, x) \mapsto g.x
$$

so that $\text{id}.x = x$ for all $x \in X$ and $g_1(g_2.x) = (g_1g_2).x$ for $g_1, g_2 \in G$, $x \in X$. We are going to drop the dot and denote actions by concatenation, $gx$. Note: this is technically a left action; a right action would flip $g_1$ and $g_2$ in the defining property.

Example 1. $X = G$, action by group multiplication. That is, $gx$ interpreted as multiplication in $G$.

Example 2. $G = \text{SO}(n) = \{g \in \text{SL}(n, \mathbb{R}) : g^{-1} = g^T\}$, $X = \mathbb{R}^n$. This is the action on $\mathbb{R}^n$ by rotations. Here $x$ is a column vector, and $gx$ is multiplication of an $n \times n$ matrix by an $n \times 1$ vector.

Example 3. $X$ is the set of bracelets of $n$ beads (placed at vertices of a regular $n$-gon) of $k$ different colours, $G$ is the dihedral group of a regular $n$-gon. Burnside’s Lemma helps count the number of such bracelets as a function of $n$ and $k$. This example illustrates that actions are interesting even when $X$ is merely a set.

An isometry is a map $\iota : \mathbb{H}^2 \to \mathbb{H}^2$ such that $d(z_1, z_2) = d(\iota(z_1), \iota(z_2))$. That is, isometries are those maps that preserve distance. A group $G$ acts by isometries on $\mathbb{H}^2$ if for every $g \in G$, the map $z \mapsto gz$ is an isometry.
We now define an action that will turn out to be an action by isometries. Let $\mathbb{CP}^1$ denote the complex projective line; it is the set of lines through the origin in $\mathbb{C}^2$, defined rigorously as

$$\mathbb{CP}^1 = (\mathbb{C}^2 \setminus \{0\})/\sim,$$

where $(z_1, z_2) \sim (z_3, z_4)$ when $(z_1, z_2) = z_5(z_3, z_4)$ for some $z_5 \in \mathbb{C} \setminus \{0\}$. The topology is the quotient topology, and we denote points by $[z_1, z_2]$. The equivalence relation $\sim$ identifies (nonzero) multiples, which are precisely points that lie on a line through the origin. We can choose preferred representatives for points of $\mathbb{CP}^1$ as follows. If $z_2 \neq 0$, then $[z_1, z_2] = [z_1/z_2, 1] = [z_3, 1]$ for some $z_3 \in \mathbb{C}$. If $z_2 = 0$ on the other hand, then $z_1 \neq 0$ since $z_1$ and $z_2$ cannot vanish simultaneously. So, $[z_1, 0] = [1, 0]$. Combining the two options, we get the identification

$$\mathbb{CP}^1 = \{[z, 1], z \in \mathbb{C}\} \cup \{[1, 0]\}.$$

The “extra” point $[1, 0]$ should be thought of as the “point at infinity,” while $[z, 1]$ is an embedded copy of $\mathbb{C}$ and therefore also $\mathbb{H}^2$, which is what we are looking for. Let

$$G = \text{GL}(2, \mathbb{R}) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \text{Mat}_{2 \times 2}(\mathbb{R}) : ad - bc \neq 0 \right\};$$

this is called the general linear group. This multiplicative group consists of invertible $2 \times 2$ matrices with real entries; note that $ad - bc = \text{det} \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. The group $G$ acts on $\mathbb{C}^2$ by matrix multiplication:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \begin{pmatrix} az_1 + bz_2 \\ cz_1 + dz_2 \end{pmatrix}.$$

This action projects to $\mathbb{CP}^1$ to give

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} z \\ 1 \end{pmatrix} = \begin{pmatrix} az + b \\ cz + d \end{pmatrix} = \begin{pmatrix} \frac{az + b}{cz + d} \\ 1 \end{pmatrix}.$$

and

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

Here $\begin{pmatrix} z \\ 1 \end{pmatrix}$ should be interpreted as $[z, 1]$, with the actual action coming from matrix multiplication.

**Theorem 2.** The group $G^+ = \text{GL}^+(2, \mathbb{R}) = \{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \text{Mat}_{2 \times 2}(\mathbb{R}) : ad - bc > 0 \} < G$ gives an action on $\mathbb{H}^2$ as defined above.

**Proof.** By $\mathbb{H}^2$ we mean $\{[z, 1] : \text{Im } z > 0\}$. Then, by definition, the equation is $z \mapsto \frac{az + b}{cz + d}$. We need to confirm that $cz + d \neq 0$ and that the imaginary part of $\frac{az + b}{cz + d}$ is positive. For the first one, assume $c \neq 0$. Then, $cz + d = 0$ implies that $z = -d/c \in \mathbb{R}$, contradicting the fact that $z$ has positive imaginary part. Now assume $c = 0$. Then, $cz + d = 0$ implies that $d = 0$, too, whence $\text{det} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = 0 \neq 1$. 


For the second part, we compute

\[
\text{Im} \frac{az + b}{cz + d} = \text{Im} \frac{ac|z|^2 + bd + adz + bc\bar{z}}{c^2|z|^2 + d^2} = \frac{\text{Im}((ad - bc)iy)}{c^2|z|^2 + d^2} = \frac{(ad - bc)y}{c^2|z|^2 + d^2}.
\]

By assumption \(ad - bc > 0\) and \(y > 0\), so the new imaginary part is positive.

Thus, the action of \(G^+\) on \(\mathbb{CP}^1\) preserves \(\mathbb{H}^2\), and in particular defines an action on \(\mathbb{H}^2\).

The transformation \((a \ b \ c \ d)^z = \frac{az + b}{cz + d}\) is called a Möbius transformation or a fractional linear transformation. We need to make this action more efficient (not a term). Observe that multiples of the identity matrix don’t move points at all: \(z \mapsto \frac{\lambda z + \beta}{\alpha z + \gamma} = z\). We remove these redundancies by reducing the group \(G^+\) a bit. Firstly, assume that the determinant of every matrix in the group is 1. This gives the special linear group

\[
\text{SL}(2, \mathbb{R}) = \{g \in G^+ : \det g = 1\}.
\]

This is indeed a group under multiplication since determinant is multiplicative. We can optimise further since the matrix \(\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}\) acts trivially. To this end define the projective special linear group

\[
\text{PSL}(2, \mathbb{R}) = \text{SL}(2, \mathbb{R})/\{\pm I\},
\]

where \(I\) is the identity matrix. Each element of the latter group corresponds to an element of \(\text{SL}(2, \mathbb{R})\) together with its negative (all entries negated).

**Theorem 3.** The group \(\text{PSL}(2, \mathbb{R})\) acts by isometries on \(\mathbb{H}^2\). In fact, it preserves the length of any curve and the area of any region.

**Proof.** Direct (and perhaps somewhat unpleasant) computation. \(\square\)

We analyse several classes of elements of \(\text{PSL}(2, \mathbb{R})\). Let

\[
r(\theta) = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix};
\]

this matrix corresponds to rotation on \(\mathbb{R}^2\), and one can check directly that \(r(\theta)i = i\). But it has another important property: no other matrix fixes \(i\). Writing

\[
\text{Stab}_G(z) = \{g \in G : gz = z \text{ for all } z \in \mathbb{H}^2\}
\]

for the stabiliser of \(x\) in \(G\), we claim that \(\text{Stab}_G(i) = \{r(\theta), \theta \in [0, \pi]\}\). Well, since \(\frac{a+i+b}{c+i+d} = i\), we cross-multiply to get \(ai + b = -c + di\). Since \(a, b, c, d\) are real, we conclude that \(a = d\) and \(c = -b\). Together with the condition that \(ad - bc = 1\), these give \(a^2 + b^2 = 1\), meaning that we can write \(a = \cos \theta\) and \(c = \sin \theta\). The reason for \(\theta \in [0, \pi]\) rather than \([0, 2\pi]\) is that the range \([\pi, 2\pi]\) is already included since in \(\text{PSL}(2, \mathbb{R})\) every element is paired with its negative.
This fact has some useful corollaries. Write
\[
n(x) = \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} \quad a(y) = \begin{pmatrix} \sqrt{y} & 0 \\ 0 & 1/\sqrt{y} \end{pmatrix}
\]
for \(x \in \mathbb{R}\) and \(y > 0\). Below are two decompositions for the Lie group \(\text{PSL}(2, \mathbb{R})\); they exist for other Lie groups.

**Theorem 4** (Iwasawa, or NAK decomposition). Every \(g \in \text{PSL}(2, \mathbb{R})\) can be written as
\[
g = n(x)a(y)r(\theta)
\]
for suitable \(x, y, \theta\).

**Proof.** We can certainly write \(g = x + iy \in \mathbb{H}^2\) for some \(x, y\). Now observe that \((n(x)a(y))^{-1}g = i\).

Thus, \((n(x)a(y))^{-1}g \in \text{Stab}_{\mathbb{G}}(i)\), and by the previous theorem there exists \(\theta\) such that \((n(x)a(y))^{-1} = r(\theta)\), as needed. \(\square\)

With some more care one can show that Iwasawa decomposition is **unique**, which is not true of the decomposition in the next theorem.

**Theorem 5** (polar, or Cartan, \(\text{KA}^+K\) decomposition). Every \(g \in \text{PSL}(2, \mathbb{R})\) can be written as
\[
g = r(\theta_1)a(y)r(\theta_2)
\]
for suitable \(\theta_1, \theta_2, y\).

**Proof.** Done in class, similar to previous theorem. \(A^+\) in the name refers to the fact that \(y\) can be chosen to be \(\geq 1\). \(\square\)

Note that if \(g = n(x)a(y)r(\theta)\) in the Iwasawa decomposition, then \(g = r(\theta_1)a(y')r(\theta_2)\) for different and unrelated \(\theta_1, y', \theta_2\).

In order to better understand the action of \(\text{PSL}(2, \mathbb{R})\) on \(\mathbb{H}^2\) (especially of the elements \(n(x)\) and \(a(y)\)) we introduce the **boundary**, or the **boundary at infinity**, of \(\mathbb{H}^2\). It is defined as
\[
\partial\mathbb{H}^2 = \{y = 0\} \cap \{\infty\}.
\]
Note that it is **not** part of \(\mathbb{H}^2\), nor is it the boundary of \(\mathbb{H}^2\) as a subset of \(\mathbb{R}^2\), but it can be thought of as the topological boundary of \(\mathbb{H}^2 \subset \mathbb{C}P^1\). The action of \(\text{PSL}(2, \mathbb{R})\) on \(\hat{\mathbb{H}}^2 = \mathbb{H}^2 \cup \partial\mathbb{H}^2\) comes from the action on \(\mathbb{C}P^1\): we use the same fractional linear transformation. That is, for any \(z \in \hat{\mathbb{H}}^2\), set
\[
\begin{pmatrix} a & b \\ c & d \end{pmatrix} z = \frac{az + b}{cz + d},
\]
with the following caveats. If \(cz + d = 0\), then interpret the fraction as \(\infty\) (exercise: \(az + b\) and \(cz + d\) do not vanish simultaneously so this is sensible). If \(z = \infty\), interpret the result as \(\frac{a}{c}\) when \(c\) is nonzero and as \(\infty\) otherwise. Once again, \(a\) and \(c\) do not vanish simultaneously.

The utility of the boundary, which is infinitely far away in the hyperbolic metric from any point of \(\mathbb{H}^2\), is it transforms together with \(\mathbb{H}^2\) under isometries. For example, we saw that the vertical line \(I\) through \(i\) is a geodesic (in fact, any vertical line is a geodesic). It can be thought of as joining \(0 \in \partial\mathbb{H}^2\) to \(\infty \in \partial\mathbb{H}^2\). In particular, for any \(g \in \text{PSL}(2, \mathbb{R})\), \(g(I)\) is again a geodesic (since \(g\) is an isometry), but moreover it joins \(g(0)\) to \(g(\infty)\).
Theorem 6. Every geodesic in $\mathbb{H}^2$ is either a vertical line or a (Euclidean) semicircle normal to the boundary.

Proof. We already saw that vertical lines are geodesics. Let $I = \{iy, y > 0\}$. Every geodesic is of the form $g(I)$ for some $g$ (why?). Put $g = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$. Then, $g(\infty) = a/c$ and $g(0) = b/d$. It is enough to prove that

$$\left| \frac{ay + b}{cy + d} - \frac{a/c + b/d}{2} \right| = \left| \frac{a/c - b/d}{2} \right|$$

for all $y$. This follows by direct computation.

In fact we can classify all isometries of $\mathbb{H}^2$. Observe that $z \mapsto -\bar{z}$ (aka $x + iy \mapsto -x + iy$) is an isometry. Geometrically it is a reflection in the $y$-axis, and it is different from any isometry in $\text{PSL}(2, \mathbb{R})$ in that it reverses orientation. This means that any circle in $\mathbb{H}^2$ traversed clockwise will be traversed anticlockwise after applying this map.

Theorem 7. The group of isometries of $\mathbb{H}^2$ is generated by Möbius transformations coming from $\text{PSL}(2, \mathbb{R})$ and the map $z \mapsto -\bar{z}$. The only orientation-preserving isometries are Möbius transformations.

Sketch. Let $\iota$ be an isometry. Then, the image of the geodesic $I = \{iy, y > 0\}$ under $\iota$ is another geodesic. We have seen in class that any pair of (distinct) points on the boundary can be mapped to any other (distinct) pair by an element of $\text{PSL}(2, \mathbb{R})$ and that geodesics joining a pair of points are mapped to geodesics joining the image points. So, let $g$ be an element of $\text{PSL}(2, \mathbb{R})$ that maps $\iota(I)$ to the geodesic joining 0 to $\infty$.

We have that $g\iota(I) = I$, but is it true that $g\iota(iy) = iy$ for every $y > 0$? Well, not necessarily, but there is no loss of generality in assuming that this is the case. Indeed, say $g\iota(i) = iY$ for some $Y \neq 1$, then we replace $g$ by $a(1/\sqrt{Y})g \in \text{PSL}(2, \mathbb{R})$, and there is no loss of generality in assuming that $g\iota(i) = i$. Since $I$ is a geodesic and both $g$ and $\iota$ are isometries, the only other thing that could “go wrong” is that $g\iota(iy) = iy$; that is, the geodesic is reversed, or, using our remarks on the boundary, the endpoints 0 and $\infty$ are interchanged. Then, replace $g$ by $r(\pi/2)g$, and we have $g\iota(iy) = iy$ for every $y > 0$.

Then, use the formula $2 \cosh d(gi, i) = a^2 + b^2 + c^2 + d^2$ to conclude that $g\iota$ is either the identity or the map $z \mapsto -\bar{z}$.

With the help of geodesics, we were able to prove that the isometry group of the hyperbolic space is precisely $\text{PSL}(2, \mathbb{R})$. Now we can classify different isometries based on matrices they come from by using the absolute value of the trace. Isometry $g$ is elliptic, parabolic, or hyperbolic according as $|\text{tr} g|$ is less that 2, equal to 2, or greater than 2. The identity matrix, which has trace 2, should be excluded from this classification.

A prime example of an elliptic isometry is $r(\theta)$, with $|\text{tr} r(\theta)| = 2|\cos \theta| < 2$ unless $\theta = n\pi$ for some $n \in \mathbb{Z}$. Recall that $r(\theta)$ is just rotation about the fixed point $i$. Observe that any conjugate of $r(\theta)$ is also elliptic (prove this) and that it is nothing more than a rotation about some other point (prove this, too).

For parabolic elements, the prototypical example is $n(x)$. Again, its conjugates have the same trace, so they, too, are parabolic.
Examples of hyperbolic elements come from the last remaining factor of the Iwasawa decomposition, \( a(y) \). They can of course come in other forms, such as, say,

\[
\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} 34 & 55 \\ 55 & 89 \end{pmatrix}.
\]

[does the last matrix look familiar?]

What does the absolute value of the trace mean geometrically? Well, we need a definition:

\( z \in \hat{H}^2 \) is a fixed point of \( g \) if \( gz = z \).

**Theorem 8.** An element of \( g = \text{PSL}(2, \mathbb{R}) \) is elliptic, parabolic, or hyperbolic according as the number of fixed points of \( g \) on the boundary \( \partial \mathbb{H}^2 \) is zero, one, or two.

The proof is an exercise. For example, note that \( r(\theta)i = i \), but \( i \) is not on the boundary, and that in fact \( r(\theta) \) moves (rotates) the boundary. On the other hand, \( n(x) \) fixes \( \infty \), only, while \( a(y) \) fixes 0 and \( \infty \).

The last group of curves we define correspond to \( n(x) \) and are called horocycles. They are of the form

\[ g \{ x + i : x \in \mathbb{R} \} \]

for some \( g \in \text{PSL}(2, \mathbb{R}) \), and the horizontal line \( \{ x + i : x \in \mathbb{R} \} \) is the basic example. As we shall see, it might be useful to include the point \( \infty \in \partial \mathbb{H}^2 \) in this horocycle.

**Theorem 9.** Horocycles in \( \mathbb{H}^2 \) are horizontal lines and Euclidean circles tangent to the boundary.

**Proof.** Let \( g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \), as usual. Since we wish to show that a certain curve is a circle, we need candidates for the centre and the radius. Following the advice to include \( \infty \) in the horocycle, we note that

\[ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \infty = \frac{a}{c}, \]

so the horocycle \( g \{ x + i \} \) meets the boundary at \( a/c \) (if \( c = 0 \), check that we in fact get a horizontal line).

Now we need another point on the horocycle. A convenient choice is the highest point (largest \( y \)-coordinate). We have

\[ \text{Im} \left( \begin{pmatrix} a & b \\ c & d \end{pmatrix} (x + i) \right) = \frac{1}{(cx + d)^2 + c^2}, \]

which should be interpreted as a function of \( x \) with \( c \) and \( d \) fixed. As \( x \to \pm \infty \), the above expression approaches 0 (assuming \( c \neq 0 \)), which is good: we expect \( \left( \frac{a}{c} \right)(x + i) \) to approach \( a/c \in \partial \mathbb{H}^2 \). The highest point is achieved when \( cx + d \) vanishes, which happens at \( x = -d/c \), and the value is \( 1/c^2 \).

Thus, we need to verify that for all \( x \in \mathbb{R} \),

\[ \left| \frac{a(x + i) + b}{c(x + i) + d} - \left( \frac{a}{c} + \frac{i}{2c^2} \right) \right| = \frac{1}{2c^2}. \]

This is an exercise. \( \square \)
Horocycles have the following geometric interpretation. Fix a point \( p \) on the boundary, and consider the family of horocycles based at \( p \). Also consider the family of geodesics joining \( p \) to another point on the boundary. Then, each horocycle is normal to each geodesic. Indeed, such a horocycle and a geodesic intersect at two points, one on the boundary and one in \( \mathbb{H}^2 \). They are clearly orthogonal at the boundary intersection. Since two (Euclidean) circles are symmetric about the line joining their centres, they must be orthogonal at the other intersection point. Alternatively, note that the “favourite” horocycle \( \{x + i\} \) is orthogonal to every geodesic of the from \( \{X + iy : y > 0\} \). Since isometries preserve angles, orthogonality also holds for isometric images.

An important observation is that in \( \mathbb{H}^2 \), unlike in \( \mathbb{R}^2 \), a horocycle is never a geodesic. That is, it never realises the minimal distance between points.

Just like in Euclidean space, in hyperbolic space there is a way to find areas of triangles based on some data about the triangle. By a triangle we mean the region bounded by three geodesics, assuming that any two of the three geodesics meet at a point in \( \hat{\mathbb{H}}^2 \). Given two smooth intersecting curves in \( \hat{\mathbb{H}}^2 \), the angle of intersection between them is the angle their tangent lines make at the point of intersection. For example, when two geodesics meet at the boundary, their angle of intersection is necessarily zero.

**Theorem 10.** Let \( T \) be a triangle in \( \hat{\mathbb{H}}^2 \) with angles \( \alpha \), \( \beta \), and \( \gamma \). Then, its area is

\[
\pi - \alpha - \beta - \gamma.
\]

**Proof.** First, suppose that \( T \) has at least one vertex on the boundary. Then, we apply an isometry so that this vertex is mapped to infinity. The resulting triangle (call it \( T \) again) is bounded by two vertical geodesics and a third semicircular geodesic. Apply an isometry (translation followed by \( z \mapsto \lambda z \) for an appropriate \( \lambda \)) so that the third geodesic is mapped to \( \{x^2 + y^2 = 1, y > 0\} \).

Thus after applying several isometries we end up with a triangle (call it \( T \) yet again) bounded by two vertical geodesics and \( \{x^2 + y^2 = 1, y > 0\} \). It is (more) convenient to compute the area by using the volume element now. Assume \( \alpha \) and \( \beta \) are angles at the semicircular geodesic, vertex corresponding to \( \alpha \) to the left of that corresponding to \( \beta \). Then, by elementary trigonometry

\[
T = \{- \cos \alpha < x < \cos \beta, y > \sqrt{1-x^2}\}.
\]

The area of \( T \) is

\[
A(T) = \int_{x=\cos \alpha}^{\cos \beta} \int_{y=\sqrt{1-x^2}}^\infty \frac{dy}{y^2} \frac{dx}{x} = \int_{x=-\cos \alpha}^{\cos \beta} \frac{dx}{\sqrt{1-x^2}} = \arcsin \cos \beta - \arcsin(-\cos \alpha) = \frac{\pi}{2} - \beta + \frac{\pi}{2} - \alpha = \pi - \alpha - \beta,
\]

as needed, since \( \gamma = 0 \). (If you don’t recall the formula for the integral, use substitution \( x = \cos u \).)
The case when $T$ does not have a vertex at infinity is an exercise (critically using the previous case!).

Example 1. Any triangle having all three of its vertices at infinity is known as the ideal triangle. Its area is clearly $\pi$. Question: are all ideal triangles isometric?

Example 2. Half of an ideal triangle (bounded say by $\{x^2 + y^2 = 1, y > 0\}$, $x = 1$, and $x = 0$) has angles $\pi/2$, 0, and 0, and hence an area of $\pi/2$, as expected.

The last topic we touch on is discrete subgroups of $\text{PSL}(2, \mathbb{R})$. The example for this course is $\text{PSL}(2, \mathbb{Z})$, which is the group of integer determinant one matrices under multiplication (with the same identification of a matrix with its negative as in the real case). Note that this set does indeed form a group (why?).

Exercise: prove that $\text{PSL}(2, \mathbb{Z})$ is a discrete subgroup of $\text{PSL}(2, \mathbb{R})$. That is, prove that the identity is an isolated point.

An example where a similar phenomenon occurs is $\mathbb{Z}$ as a subgroup of $\mathbb{R}$, or indeed $\mathbb{Z}^n < \mathbb{R}^n$. We certainly have that 0 is an isolated point in $\mathbb{Z}^n$.

It is proved in Einsiedler and Ward (copies I gave out) that, as a subgroup of the isometry group of $\mathbb{H}^2$, it is generated by $\tau: z \mapsto z + 1$ and $\sigma: z \mapsto -1/z$. Recall that $\tau$ translates to the right by 1, while $\sigma$ “flips” about the geodesic joining 0 to $\infty$. Alternatively, one can think of $\sigma$ as the Möbius transformation corresponding to $r(\pi/2)$, which is a rotation by $\pi$ about $i$.

An important notion for discrete subgroups of $\text{PSL}(2, \mathbb{R})$ is that of a fundamental domain: $\mathcal{F} \subset \mathbb{H}^2$ is a fundamental domain for a discrete group $\Gamma$ if the area of $\mathcal{F} \cap \gamma \mathcal{F}$ is zero for $\gamma \neq 0$ and the area of

$$\mathbb{H}^2 \setminus \bigcup_{\gamma \in \Gamma} \gamma \mathcal{F}$$

is zero. In the strict sense, one can require that the sets be empty instead of zero area, but the definition above is more convenient in our situation. Effectively it allows us to ignore boundaries of regions. A group $\Gamma$ is call a lattice if there exists a finite area fundamental domain for $\Gamma$.

Here is a concrete and somewhat analogous real example. Consider $\Gamma = \mathbb{Z} < \mathbb{R} = G$ under addition. Then, $[0, 1)$ is fundamental domain for $\Gamma$ in the strict sense. For the first condition, observe that $[0, 1)$ and $n + [0, 1)$ are disjoint for every nonzero integer $n$. For the second, note that the union $\bigcup_{n \in \mathbb{Z}} (n + [0, 1))$ does contain the entire real line. Other choices of strict fundamental domains are $(0, 1]$, $(-3, -2]$, etc. Examples of fundamental domains as defined above are $[0, 1]$, $(-4, -3)$, etc., with points at the boundary being irrelevant from the point of view of length rather than area.

There is never a unique fundamental domain, but their may be a preferred one. In the real example, $[0, 1)$ is the standard choice, which corresponds to mapping $x \in \mathbb{R}$ to the fractional part of $x$ in $[0, 1)$. Another familiar example where fundamental domains are relevant includes integers modulo $n$, where the preferred choice is $\{0, \ldots, n - 1\}$. One may equally well choose $\{0, n + 1, 2n + 2, 8n + 3, n^{51} + 4, \ldots, -1\}$, which would be quite non(sub)standard.

How do we construct a nice fundamental domain for $\text{PSL}(2, \mathbb{Z})$? We use the generators $\sigma$ and $\tau$ as defined above. Since $\tau$ acts by translation by 1, there is a fundamental domain within any vertical strip of unit width.
Suppose for contradiction that there isn’t a fundamental domain in a strip but that a fundamental domain $F$ exists (an ugly fundamental domain always exists by the axiom of choice). Let $F_n = ([n, n+1) \times \mathbb{R}_{>0}) \cap F$. Then, $F' = \cup_{n \in \mathbb{Z}} \tau^{-n} F_n$ is an almost disjoint union (elements may only overlap in a set of zero area) and is a fundamental domain. Both conditions are satisfied because $F$ is a fundamental domain (check this). However, $F'$ is contained in the strip $[0,1) \times \mathbb{R}_{>0}$, contradicting our assumption.

The same argument applies more generally. The map $\sigma$ switches the interior of the (semi)circle $\{x^2 + y^2 = 1\}$ with its exterior; therefore, either side contains a full fundamental domain. A careful argument (cf. Einsiedler and Ward) confirms that the set

$$\{x^2 + y^2 > 1\} \cap \{|x| < 1/2\} = F$$

constitutes a fundamental domain for $\text{PSL}(2,\mathbb{Z})$. This is the standard fundamental domain for this group. Other options include $\gamma F$ for any $\gamma \in \text{PSL}(2,\mathbb{Z})$ or, say, take any partition $F = F_1 \cup F_2$ and use $\gamma_1 F_1 \cup \gamma_2 F_2$. Do have a look at E&W for pictures.

This set has many (I mean, MANY) interesting connections to number theory, dynamical systems, complex analysis, representation theory, knot theory, geometry of numbers, and so on.