Introduction to hyperbolic geometry (PART I)

Binbin Xu

August 29, 2021

Contents

1 Preliminary

This note is based on a mini-course given in Chern Institute, Nankai University, in July 2021.
Pre-required Knowledge:

- Linear Algebra (2×2 matrices with real or complex coefficients);
- Analysis (limits, Riemann integral, (partial) derivative of real value functions);
- Geometry in Euclidean plane (line, circle, length, angle, tangent, computation with different coordinate systems)
- Basic notions in Group Theory (group, subgroup, generator, relations, presentations of groups, homomorphism, isomorphism)
- Basic notions in Topology (basis, open/closed set, fundamental group, universal cover)

1.1 Introduction

Euclidean geometry is one geometry (maybe **the** geometry) with which we are familiar the most. Its axiomatic system is based on 5 postulates where the 5th one which is usually called the 'Parallel Postulate'. An equivalent statement of this postulate given by Playfair says that:

'In a plane, through a point not on a given straight line,
at most one line can be drawn that never meets the given line.'

By modifying this postulate, we are getting into the world of non-Euclidean geometry which includes two main types: spherical geometry and hyperbolic geometry. As one main type of the non-Euclidean geometry, hyperbolic geometry not only is a beautiful and rich research area of mathematics by itself, but also has connections to various other areas in mathematics and physics, such as dynamical system, geometric group theory, number theory, projective geometry, mathematical physics, etc.

The 2-dimensional hyperbolic geometry is the basic of this area. The main objects studied in this area are hyperbolic surfaces which, on one hand, admit many interesting properties and a rich deformation theory, and at the same time, provide elementary examples in different research areas. In this mini course, we would like to give an introduction to this topic.

1.2 First impression of non Euclidean geometry

Before getting into details, we may start by comparing triangles and disks in non-Euclidean space with those in the Euclidean plane to get a first impression of non-Euclidean geometry.

Recall that in the Euclidean plane, any triangle have the sum of interior angles to be π. Now let us consider the Earth. Assume that it is a ball. Two persons A and B walk from the north pole to the equator along different meridians. Then when they get to the equator, their traces and the segment in the equator connecting them form a triangle in the sphere. Notice that there are two angles which are already right angles. Hence the sum of interior this triangle on the sphere is bigger than π.

In fact, this is a general phenomenon: in the spherical geometry, the sum of interior angles of a triangle is bigger than π, and the triangle looks "fatter" than a triangle with the same side lengths in the Euclidean plane.

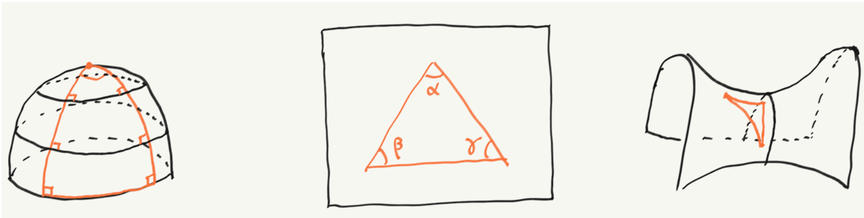

Figure 1.2.1: $\alpha + \beta + \gamma$: Spherical > Euclidean$= \pi >$ Hyperbolic

Then we may wonder: what about the triangles in the hyperbolic geometry? Since we have "strictly bigger than π", and "equal to π", one possible guess would be: in hyperbolic geometry, the sum of interior angles of a triangle is strictly smaller than π. This is exactly what happens. Moreover, a triangle in the hyperbolic space looks "thinner" than the triangle with the same side lengths in the Euclidean plane. We will explain these in this mini-course.

Another object which we meet a lot when studying Euclidean geometry is the disk. We may take a piece of paper and cut a disk out. Then we can try to cover some part of a sphere (one may think about covering a chocolate ball). No matter how we adjust the paper disk, there is always some part of the disk folded and the disk overlaps with itself. This means that we have some extra area, and the area of the paper disk is bigger than that of the disk that it covers on the sphere.

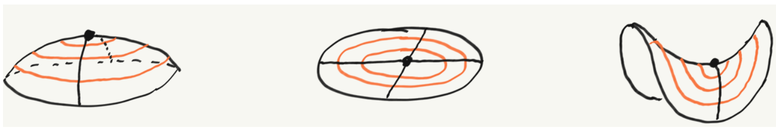

Figure 1.2.2: Areas of disks of a same radius: Spherical < Euclidean < Hyperbolic

On the other hand, if we try to use the paper disk to cover the top surface of a saddle, we may find that if we push the paper disk to make it touching the saddle, then the paper breaks along radius. This means that there is not enough area. In the other words, the area of the paper disk is smaller than that of the disk that it tries to cover on the saddle. We will explain this phenomenon later in this mini-course as well.

1.3 Plan of the mini course

In this mini course, we will focus on the hyperbolic geometry, more precisely, 2-dimensional hyperbolic geometry. The mini course contains three parts:

(i) In the first part, we give an elementary introduction to the geometry of the hyperbolic plane by discussing lines, circles and triangles in this space. We will also talk about the isometry group of the hyperbolic plane and its discrete subgroups.

(ii) In the second part, we will study the geometry on hyperbolic surfaces. We will end this part by introducing briefly the Teichmüller space and the mapping class group which are the main objects studied in the Teichmüller theory and are closely related to the study of hyperbolic surfaces.

(iii) In the end of the mini course, we will briefly discuss two interesting topics in the study of hyperbolic surfaces: identities associated to hyperbolic surfaces and counting curves on hyperbolic surfaces.

1.4 References

There are many references for studying hyperbolic geometry and hyperbolic surfaces. Here is a short list of references which I use a lot.

[1] Alan F. Beardon, The geometry of discrete groups, *Graduate Texts in Mathematics, vol. 91, Springer-Verlag, New York, 1983.*

[2] Benson Farb and Dan Margalit, A primer on mapping class groups, *Princeton Mathematical Series, vol. 49, Princeton University Press, Princeton, NJ, 2012.*

[3] Travaux de Thurston sur les surfaces, *Séminaire Orsay, Astérisque No. 66-67, Société Mathématique de France, Paris, 1991.*

[4] Svetlana Katok, Fuchsian groups, *Chicago Lectures in Mathematics, University of Chicago Press, Chicago, IL, 1992.*

2 Hyperbolic plane

2.1 Models of hyperbolic plane

Instead of using the abstract definition, we usually study hyperbolic space using its different models. Below is a list of four models that we often meet:

- Upper half plane model;
- Poincaré disk model;
- Kleinian model;
- Minkowski model.

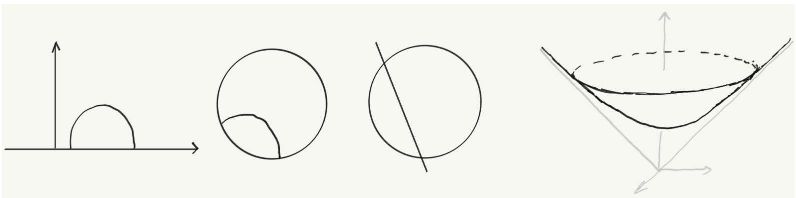

Figure 2.1.1: Upper half plane model; Poincaré disk model; Kleinian model; Minkowski model

Different models have different advantages. We usually choose the model according to the problem that we would like to study, or maybe use several models at the same time. The most basic models are the first two which we will introduce in the following.

2.2 Upper half plane model $\mathbb{H}$

As described in its name, the set for this model is the upper half of the complex plane $\mathbb{C}$:

$$\mathbb{H} = \{x + iy \in \mathbb{C} \mid y > 0\}.$$

The hyperbolic metric (length element) can be described using coordinates (x, y) as follows:

$$\mathrm{d}s_{\mathbb{H}} = \frac{\sqrt{\mathrm{d}x^2 + \mathrm{d}y^2}}{y},$$

which is given by renormalizing the Euclidean metric by the y coordinate (See Figure 2.2.1).

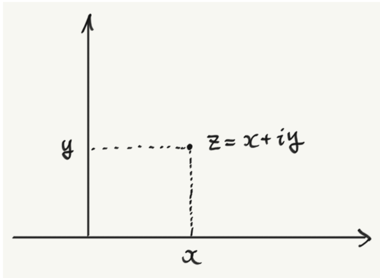

Figure 2.2.1: $\mathbb{H}$ with Cartesian coordinates (for Euclidean plane)

Informally speaking, one may think the upper half plane is cut into very thin horizontal strips. On each horizontal strip, the metric is Euclidean metric, but renormalized by the the inverse of the y-coordinate of a boundary of the strip. Hence if we have a segment, under the Euclidean metric its length is independent of its position. Now under the hyperbolic metric, the higher it is, the shorter it is (See Figure 2.2.2).

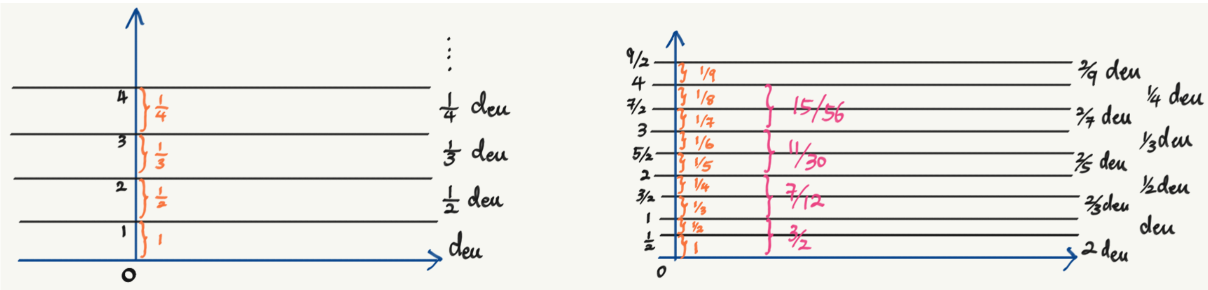

Figure 2.2.2: partition by 1; partition by $\frac{1}{2}$ (d_{eu} stands for the Euclidean length element)

On may notice that the shortest path may be changed under this new metric. If a path goes vertically, it seems still be shortest. But if a path goes horizontally, there maybe a way to find a shorter path. (See Figure 2.2.3) For example, we may go up a little then move horizontally then come down. Although going up and coming down add some distance, but the horizontal path with bigger y coordinate is shorter. In fact this is what would happen to the shortest path between points, i.e. geodesics. We will discuss in details in the next part.

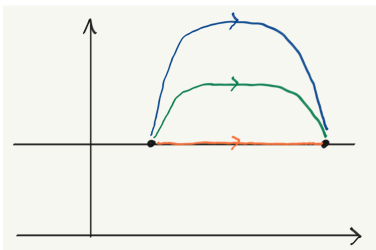

Figure 2.2.3: Which one is the shortest?

Remark 2.2.1.
If we consider the light travels in some material, we know that its speed depends on the material. We may consider the upper half plane is obtained by gluing layers of different materials together, such that the higher the material is, the faster the light travels inside this material.

Consider two points in the upper half plane. Given any path connecting them, instead of the 'length' of the path, we may consider the time that the light spends when travel along this path. The path along which the light spends the least time can be considered as the 'geodesic'.

We may consider what happens when the light goes from a point in the water to a point in the air. Instead of being an Euclidean straight segment, our experience tells us that the path is bent. (Light travels faster in the air than in the water.)

From the expression of the metric, we know that infinitesimally the hyperbolic geometry is still Euclidean geometry (up to a rescaling by a constant). The notion of angle is the same as that in the Euclidean geometry. We will use them directly without any change.

Remark 2.2.2.
We usually use $\mathbb{H}^n$ to denote n-dimensional hyperbolic space. Since we only consider dimension 2, we omit the index 2, and use $\mathbb{H}$ to denote the hyperbolic plane, instead of $\mathbb{H}^2$.

Sometimes, we also consider the polar coordinates of $\mathbb{C}$ (See Figure 2.2.4). Then we have $(x, y) = (r\cos\theta, r\sin\theta)$ and

$$\mathrm{d}x = \cos\theta\,\mathrm{d}r - r\sin\theta\,\mathrm{d}\theta$$
$$\mathrm{d}y = \sin\theta\,\mathrm{d}r + r\cos\theta\,\mathrm{d}\theta$$

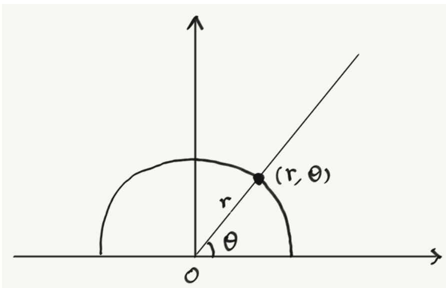

Figure 2.2.4: The point $z = re^{i\theta}$ has polar coordinates (r, θ) (for Euclidean plane)

Then the hyperbolic metric can be written as the following:

$$\mathrm{d}s_{\mathbb{H}} = \frac{\sqrt{\mathrm{d}x^2 + \mathrm{d}y^2}}{y} = \frac{\sqrt{\mathrm{d}r^2 + r^2\,\mathrm{d}\theta^2}}{r\sin\theta} = \frac{\sqrt{r^{-2}\,\mathrm{d}r^2 + \mathrm{d}\theta^2}}{\sin\theta}.$$

2.3 Distance in $\mathbb{H}$

Once we have the metric infinitesimally, we can start talking about length of paths by taking integral. Of course, in order to do so, we will use "parametrizations" and need the path to have certain regularity. More precisely, we consider the following definitions.

Definition 2.3.1
A *(parametrized) path* in $\mathbb{H}$ connecting points w and z is the image of a continuous map γ from an interval $[a,b]$ to $\mathbb{H}$ with $\gamma(a) = w$ and $\gamma(b) = z$. We call γ the *parametrization* of the path $\gamma([a,b])$.

For $t \in [a,b]$, we denote $\gamma(t) = (x(t), y(t))$ or $x(t) + iy(t)$, and each coordinate is also a continuous function on the parameter in t.

Remark 2.3.2.
We will abuse the notation and use γ to denote the corresponding path as well.

Definition 2.3.3
A path $\gamma : [a,b] \to \mathbb{H}$ is said to be *of class C^1 (or simply C^1)* if both functions

$$x : [a,b] \to \mathbb{R} \quad \text{and} \quad y : [a,b] \to \mathbb{R},$$

are of class C^1, i.e. they are differentiable and their derivative $\dot{x}$ and $\dot{y}$ are continuous.

Definition 2.3.4
The *tangent vector* of a regular path γ at $\gamma(t)$ is defined to be the vector $\dot{\gamma}(t) = (\dot{x}(t), \dot{y}(t))$.

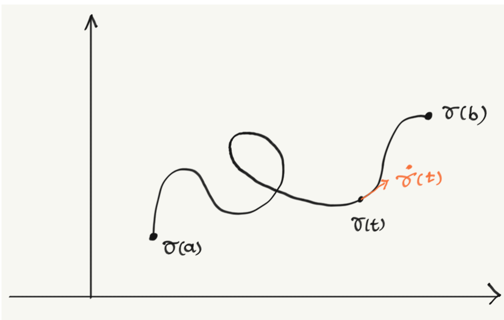

Figure 2.3.1: A regular path $\gamma : [a,b] \to \mathbb{H}$

Definition 2.3.5
A path γ is said to be *piecewise C^1* if there is a finite partition of the parameter interval $a = t_0 < t_1 < \cdots < t_n = b$ such that the restriction γ is C^1 on each part $[t_j, t_{j+1}]$ for $0 \le j \le n-1$.

We will denote its Euclidean norm by

$$|\dot{\gamma}(t)|_{\mathbb{E}} = \sqrt{\dot{x}(t)^2 + \dot{y}(t)^2}$$

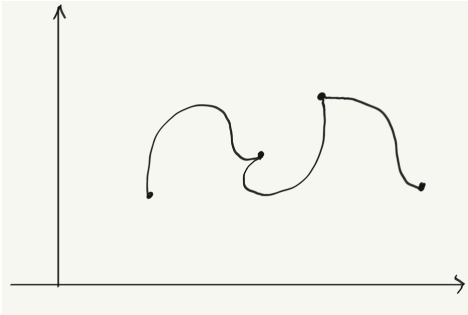

Figure 2.3.2: A piecewise regular path

and its hyperbolic norm by

$$|\dot{\gamma}(t)|_{\mathbb{H}} = \frac{|\dot{\gamma}(t)|_{\mathbb{E}}}{y(t)} = \frac{\sqrt{\dot{x}(t)^2 + \dot{y}(t)^2}}{y(t)}.$$

Let $\gamma : [a, b] \to \mathbb{H}$ be a regular path.

Definition 2.3.6
The *(hyperbolic) length* of path γ is defined to be the following quantity:

$$l_{\mathbb{H}}(\gamma) = \int_a^b |\dot{\gamma}(t)|_{\mathbb{H}} \, \mathrm{d}t = \int_a^b \frac{\sqrt{\dot{x}(t)^2 + \dot{y}(t)^2}}{y(t)} \, \mathrm{d}t.$$

Informally the above integral can be understood as follows. we may consider a partition of the regular path. At each partition point, we take the Euclidean straight line tangent to the path. By taking segments in these lines, we find a piecewise straight line. By adding more and more points to the partition to get a finer and finer partition, we may get a sequence of piecewise straight lines which approximate to the regular path. The lengths of these piecewise lines will converge to that of the regular path. By taking the limit, taking the sum becomes taking the integral.

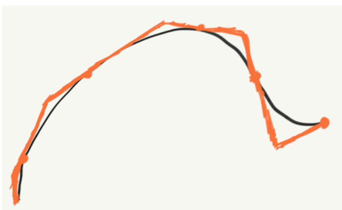

Figure 2.3.3: Approximation using piecewise Euclidean straight line

A path may have different parametrizations. However, the length of the path is independent of the parametrization.

Informally speaking, changing parameter changes only the speed used when we go through this path and how much time we spend to pass this path, but not the length of this path.

More precisely, if $\phi : s \mapsto t$ is a change of parameters which is C^1 without changing the starting and ending points. Let $\eta(s) = (\gamma \circ \phi)(s) = (u(s), v(s))$. Then we have

Proposition 2.3.7

$$l_{\mathbb{H}}(\gamma) = l_{\mathbb{H}}(\eta).$$

Proof. We verify this relation as follows:

$$\begin{aligned} l_{\mathbb{H}}(\gamma) = \int_a^b |\dot{\gamma}(t)|_{\mathbb{H}} \, \mathrm{d}t &= \int_{\phi^{-1}(a)}^{\phi^{-1}(b)} \frac{\sqrt{\dot{x}(\phi(s))^2 + \dot{y}(\phi(s))^2}}{y(\phi(s))} |\phi'(s)| \, \mathrm{d}s \\ &= \int_{\phi^{-1}(a)}^{\phi^{-1}(b)} \frac{\sqrt{\dot{u}(s)^2 + \dot{v}(s)^2}}{v(s)} \, \mathrm{d}s \\ &= \int_{\phi^{-1}(a)}^{\phi^{-1}(b)} |\dot{\eta}(s)|_{\mathbb{H}} \, \mathrm{d}s = l_{\mathbb{H}}(\eta) \end{aligned}$$

□

A different way to approximate a path is by first taking a partition of the path, then connecting the partition points using Euclidean straight segments. By taking finer and finer partition of the path, we have a sequence of Euclidean piecewise straight lines.

As before, to each Euclidean segment, we may associate to it the 'length' given by its Euclidean length renormalized by $1/y$ where y is the imaginary part of a point in this segment. In this way, to each of these Euclidean piecewise straight lines, we can associate to it a 'length'.

If the supreme of the 'length's of these Euclidean piecewise straight lines is finite, then we say that the path is *rectifiable* and its actual hyperbolic length equals to this supreme.

Remark 2.3.8.
A rectifiable path could be more general than piecewise regular paths. For example, we may construct a rectifiable path which is not differentiable on a dense subset of the parameter interval. (Bending sub-arcs in a circle arc)

In the rest of this note, most of the paths that we will consider have certain natural parametrizations. We will use them directly. Before moving further, let us see some examples.

Example 2.3.9.
Let $\gamma(t) = (0, t)$ with $t \in [a, b]$. Then $\dot{\gamma}(t) = (\dot{x}(t), \dot{y}(t)) = (0, 1)$.

Thus its length is:

$$l_{\mathbb{H}}(\gamma) = \int_a^b |\dot{\gamma}(t)|_{\mathbb{H}} \, \mathrm{d}t = \int_a^b \frac{\sqrt{0^2 + 1^2}}{t} \, \mathrm{d}t = \int_a^b \frac{1}{t} \, \mathrm{d}t = \log \frac{b}{a}.$$

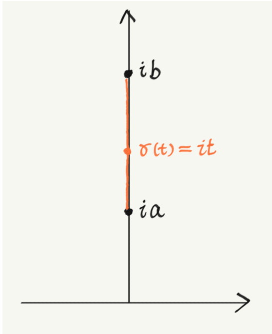

Figure 2.3.4: Path $\gamma(t) = it$ for $t \in [a, b]$

Remark 2.3.10.
The formula of the length is independent of x coordinate for the vertical line. If we consider the path $\gamma(t) = (x, t)$ with $t \in [a, b]$, the length of γ will be the same.

Example 2.3.11.
We consider the half circle centered at the origin passing i (with Euclidean radius 1). We would like to compute the hyperbolic length of the arc between e^{ia} and e^{ib} with $0 < a < b < \pi$. We use the polar coordinate of $\mathbb{C}$, and the path can be described as $\gamma(t) = (r(t), \theta(t)) = (1, t)$ for $t \in [a, b]$. Hence $\dot\gamma(t) = (0, 1)$.

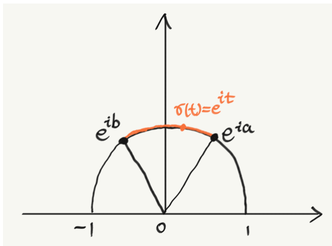

Figure 2.3.5: Path $\gamma(t) = e^{it}$ for $t \in [a, b]$

Its length is:

$$
\begin{aligned}
l_{\mathbb{H}}(\gamma) &= \int_a^b |\dot\gamma(t)|_{\mathbb{H}}\,\mathrm{d}t = \int_a^b \frac{\sqrt{0+1}}{\sin t}\,\mathrm{d}t = \int_a^b \frac{\sin^2\frac{t}{2} + \cos^2\frac{t}{2}}{2\sin\frac{t}{2}\cos\frac{t}{2}}\,\mathrm{d}t \\
&= \int_a^b \left(-\frac{\mathrm{d}\cos\frac{t}{2}}{\cos\frac{t}{2}} + \frac{\mathrm{d}\sin\frac{t}{2}}{\sin\frac{t}{2}} \right) \\
&= \log\frac{\cos\frac{a}{2}}{\cos\frac{b}{2}} + \log\frac{\sin\frac{b}{2}}{\sin\frac{a}{2}} \\
&= \log\tan\frac{b}{2} - \log\tan\frac{a}{2} \\
&= \log\frac{\sin b}{\cos b + 1} - \log\frac{\sin a}{\cos a + 1}
\end{aligned}
$$

Remark 2.3.12.
The formula of the length is independent of the center x and the radius r of the circle. If we consider the path $\gamma(t) = x + re^{it}$ with $t \in [a,b]$, the length of γ will be the same.

Definition 2.3.13
The *distance* (length metric) between w and z two points in $\mathbb{H}$ is defined to be the following quantity:

$$\mathrm{d}_{\mathbb{H}}(w,z) = \inf\{l_{\mathbb{H}}(\gamma) \mid \gamma : [a,b] \to \mathbb{H} \text{ is piecewise } C^1,\ \gamma(a) = w \text{ and } \gamma(b) = z\}.$$

Remark 2.3.14.
Alternatively, for any two points in $\mathbb{H}$, we may consider the Euclidean piecewise straight paths connecting them. By taking the infimum of the hyperbolic lengths of these paths, we get the distance between two points in $\mathbb{H}$ which is equivalent to the above one.

Proposition 2.3.15
The function $\mathrm{d}_{\mathbb{H}}$ *is a distance function.*

Proof. Let z, z' and z'' be three points in $\mathbb{H}$. It is enough to check that it satisfies all three conditions in the definition of a distance function.

(i) $\mathrm{d}_{\mathbb{H}}(z,z') \geq 0$, and $\mathrm{d}_{\mathbb{H}}(z,z') = 0$ if and only if $z = z'$;

(ii) $\mathrm{d}_{\mathbb{H}}(z,z') = \mathrm{d}_{\mathbb{H}}(z',z)$;

(iii) $\mathrm{d}_{\mathbb{H}}(z,z'') \leq \mathrm{d}_{\mathbb{H}}(z,z') + \mathrm{d}_{\mathbb{H}}(z',z'')$.

The second condition is satisfied, since a path going from z to z' can be obtained from reversing the orientation of a path going from z' to z, and verse versa.

The third condition is also satisfied. If we connect a piecewise C^1 path going from z to z' to a piecewise C^1 path going from z' to z'', we get a piecewise C^1 path going from z to z''.

The positivity part of the first condition is immediate, since the length of a path is always positive (≥ 0). To see the "if and only if" part, without loss of generality, we may consider a regular path $\gamma : [a,b] \to \mathbb{H}$ connecting z and z', and assume that $\operatorname{Im} z \leq \operatorname{Im} z'$. Since the path is compact, there is a horizontal strip in $\mathbb{H}$ contained it with boundary $y = y_0$ and $y = y_1$ with $y_0 \leq \operatorname{Im} z \leq \operatorname{Im} z' \leq y_1$.

If $y_1 \leq 2\operatorname{Im} z$, , we have

$$\begin{aligned} l_{\mathbb{H}}(\gamma) &= \int_a^b \frac{\sqrt{\dot{x}(t)^2 + \dot{y}(t)^2}}{y(t)}\,\mathrm{d}t \\ &\geq \int_a^b \frac{\sqrt{\dot{x}(t)^2 + \dot{y}(t)^2}}{2\operatorname{Im} z}\,\mathrm{d}t \\ &= \frac{l_{\mathbb{E}}(\gamma)}{2\operatorname{Im} z}. \end{aligned}$$

If $y_1 \geq 2\mathrm{Im}\, z$, then

$$\begin{aligned} l_{\mathbb{H}}(\gamma) &= \int_a^b \frac{\sqrt{\dot{x}(t)^2 + \dot{y}(t)^2}}{y(t)} \,\mathrm{d}t \\ &\geq \int_a^b \frac{\sqrt{\dot{y}(t)^2}}{y(t)} \,\mathrm{d}t \\ &= l_{\mathbb{H}}(\pi_y(\gamma)) \\ &\geq \log \frac{y_1}{y_0} \quad (y_0 \leq \Im z) \\ &\geq \log \frac{y_1}{\mathrm{Im}\, z} \quad (y_1 \geq 2\mathrm{Im}\, z) \\ &\geq \log 2. \end{aligned}$$

where π_y is the horizontal projection to the imaginary axis of $\mathbb{H}$. Hence we have

$$\mathrm{d}_{\mathbb{H}}(z, z') \geq \min \left\{ \log 2, \frac{l_{\mathbb{E}}(\gamma)}{2\mathrm{Im}\, z} \right\},$$

which is strictly bigger than 0, when z and z' are distinct. When $z = z'$, the constant path has 0 length. Therefore, we have $\mathrm{d}_{\mathbb{H}}(z, z') = 0$ if and only if $z = z'$. □

Remark 2.3.16.
In next part, we will show that for each pair of points z and w in $\mathbb{H}$, the infimum in the definition of $\mathrm{d}_{\mathbb{H}}$ is realized by a unique path whose length is 0 if and only if $w = z$. This in fact provides a constructive proof of the above lemma.

We can also define distance between non empty subsets of $\mathbb{H}$. Let K and K' be two non-empty subsets of $\mathbb{H}$.

Definition 2.3.17
The *distance* between K and K' are defined as the following quantity:

$$\mathrm{d}_{\mathbb{H}}(K, K') = \inf\{\mathrm{d}_{\mathbb{H}}(z, z') \mid z \in K,\ z' \in K'\}.$$

If moreover, there exists $z_0 \in K$ and $z_0' \in K'$, such that

$$\mathrm{d}_{\mathbb{H}}(z_0, z_0') = \mathrm{d}_{\mathbb{H}}(K, K'),$$

then we say that z_0 and z_0' *realize the distance* between K and K'.

We may check that the distance between two disjoint non empty sets is not always strictly positive. For example, two disjoint Euclidean open disks with boundaries tangent to each other have distance 0.

2.4 Geodesics in $\mathbb{H}$

Definition 2.4.1
A path $\gamma : [a, b] \to \mathbb{H}$ is said to be a *geodesic (segment)* if it locally minimizes the distance, i.e.

for any $c \in [a, b]$, there exists $\epsilon > 0$, such that for any subinterval $[t, t']$ in $[c - \epsilon, c + \epsilon] \cap [a, b]$, we have $d_{\mathbb{H}}(\gamma(t), \gamma(t')) = l_{\mathbb{H}}(\gamma|_{[t,t']})$.

We use "locally minimize" in the definition. This is because in general given two points in the space, it is possible that there are several geodesics with different lengths connecting them. See the following example on sphere:

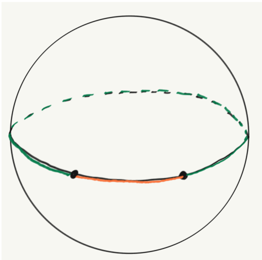

Figure 2.4.1: The two colored paths are both geodesics.

See another example with non-trivial topology:

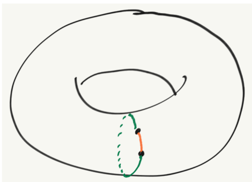

Figure 2.4.2: The two colored paths are both geodesics.

This is not the case when we consider $\mathbb{H}$. A path γ in $\mathbb{H}$ locally minimize the distance if and only if it globally minimize the distance, i.e. for any pair of parameters $a < t < t' < b$, we have $d_{\mathbb{H}}(\gamma(t), \gamma(t')) = l_{\mathbb{H}}(\gamma|_{[t,t']})$.

Proposition 2.4.2

The geodesics in $\mathbb{H}$ are either vertical lines or half circles with center at the real axis.

Proof. Without loss of generality, we consider only regular paths.
(Proof for vertical lines) Let z and w be two points on a same vertical line V defined by $x = x_0$. Let $\gamma : [a, b] \to \mathbb{H}$ be a regular arc with $\gamma(a) = z$ and $\gamma(b) = w$. We consider its horizontal projection to V and get a regular arc $\eta : [a, b] \to \mathbb{H}$. More precisely, for each $t \in [a, b]$, we have $\gamma(t) = (x(t), y(t)$ and $\eta(t) = (x_0, y(t))$.

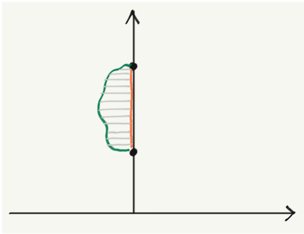

Figure 2.4.3: Paths connecting points on a same vertical line

Notice that there maybe back track in η. Then we have the following comparison:

$$\begin{aligned} l_{\mathbb{H}}(\gamma) &= \int_a^b \frac{\sqrt{\dot{x}(t)^2+\dot{y}(t)^2}}{y(t)}\,\mathrm{d}t \\ &\geq \int_a^b \frac{\sqrt{0+\dot{y}(t)^2}}{y(t)}\,\mathrm{d}t \\ &= l_{\mathbb{H}}(\eta). \end{aligned}$$

Moreover the equality is realized if and only if $\dot{x}(t)=0$ for all t, which is equivalent to $x(t)=x_0$ for all t. Hence the vertical segment connecting z and w realizes the minimum of the lengths of all piecewise C^1 path connecting them.
(Proof for half circles) The proof for the second type is similar. We use the formula for polar coordinates that we have studied previously.

Given any two points z and w which are not on a same vertical line, there is a unique half circle C which passes them with center contained in the real axis. Without loss of generality, we may consider the center of the half circle is the origin with radius r_0. (otherwise, we may consider a change of the coordinates by moving the origin to the center of the circle) Given a regular path connecting z and w, we consider its projection to C along Euclidean rays issued from the original.

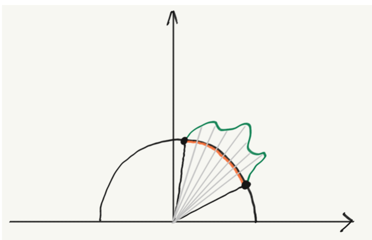

Figure 2.4.4: Paths connecting points on a same circular arc with center on the real axis

Then we have the following

$$\begin{aligned} l_{\mathbb{H}}(\gamma) &= \int_a^b \frac{\sqrt{r^{-2}\dot{r}(t)^2+\dot{\theta}(t)^2}}{\sin\theta(t)}\,\mathrm{d}t \\ &\geq \int_a^b \frac{\sqrt{0+\dot{\theta}(t)^2}}{\sin\theta(t)}\,\mathrm{d}t \\ &= l_{\mathbb{H}}(\eta). \end{aligned}$$

Moreover the equality is realized if and only if $\dot{r}(t) = 0$ for all t, which is equivalent to $r(t) = r_0$ for all t. Hence the circular arc on $C(0, r_0)$ connecting z and w realizes the minimum of the lengths of all piecewise C^1 path connecting them, where $C(0, r_0)$ stands for the Euclidean circle centered at $x = 0$ of radius r_0. □

Remark 2.4.3.
We have not only proved that the geodesic between two points is either a segment in a vertical line or an arc in a half circle with center on the real axis, but also shown that this geodesic is unique for each pair of points. Moreover, the distance between two points is 0 if and only if the geodesic is a point which means that the two points are a same one. For our convenience, we will call the former a *vertical geodesic* and the latter a *circular geodesic* of $\mathbb{H}$.

Definition 2.4.4
A *complete geodesic* is a path $\gamma : \mathbb{R} \to \mathbb{H}$ such that for any $t < t'$, we have

$$d_{\mathbb{H}}(\gamma(t), \gamma(t')) = |t - t'|.$$

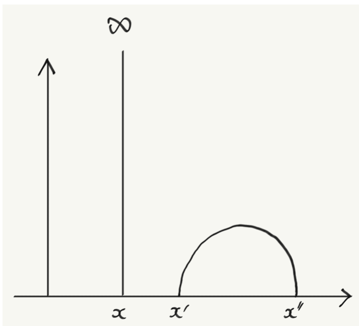

Figure 2.4.5: Complete geodesics in $\mathbb{H}$ with their end points

Since each circular geodesic is a half Euclidean circle with center on real axis, it is determined by its intersection with $\mathbb{R}$ which are two distinct points on $\mathbb{R}$. We can formally add a point ∞ to $\mathbb{R}$, then the vertical geodesic is determined by a real number and ∞. Reciprocally, given two distinct points in $\widehat{\mathbb{R}} = \mathbb{R} \cup \{\infty\}$, they determine a unique geodesic with them as ending points. Hence this induces a parametrization of the space of geodesics in $\mathbb{H}$ and which can be identified with a Möbius band. The points on $\widehat{\mathbb{R}}$ associated to each geodesic γ will be called the *end points* of γ. We will come back to this later, and will see that using $\widehat{\mathbb{R}}$ is not an arbitrary choice and that it plays an important role in the geometry of $\mathbb{H}$.

There are three relative positions between two geodesics:

- they intersect each other;
- (*Parallel*) they share one end point;
- (*Ultra-parellel*) they do not intersect each other and share no end point.

Remark 2.4.5.

Sometimes, the third type is also called disjoint (which we used in the lecture). To avoid the ambiguity when we use the word 'disjoint', we use ultra-parellel.

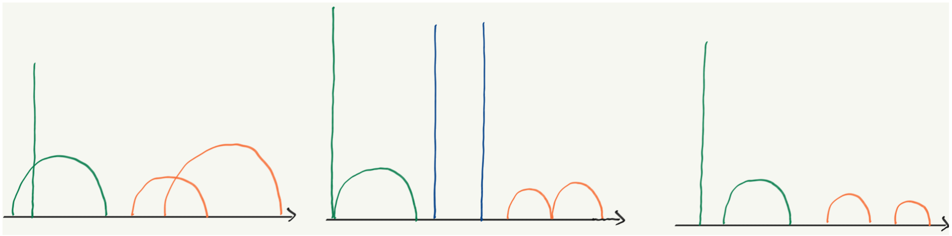

Figure 2.4.6: Relative position between geodesics

3 Distance formula and its convexity

3.1 Distance formula

Using the help of Euclidean geometry, we have described all possible types of geodesics, and we also know that any pair of points are connected by a geodesic segment. It is then possible to compute the distance formula for any pair of points, which is stated as follows.

Proposition 3.1.1

For any points w and z in $\mathbb{H}$, their hyperbolic distance is given by the following formula:

$$\mathrm{d}_{\mathbb{H}}(w, z) = \log \frac{|\overline{w} - z| + |w - z|}{|\overline{w} - z| - |w - z|}.$$

Proof. Notice that we can classify the relative position between two points into two types:

- They are on the same vertical geodesic;
- or not, then there is a unique circular geodesic passing through them.

We will try to get a formula for the distance between a pair of points of the second type, then show that it also holds for the first type.

Let w and z be two points in $\mathbb{H}$ which are not on a same vertical geodesic. There is a unique Euclidean circle C passing through them with center on the real axis. Let x be the center and r be the radius of C.

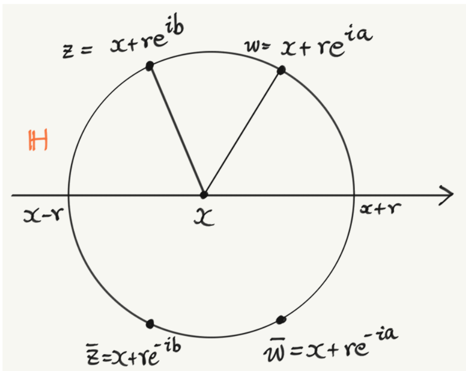

Figure 3.1.1: Setting

Then w and z can be expressed as:

$$w = x + re^{ia}$$
$$z = x + re^{ib}$$

for some a and b in $(0, \pi)$. Without loss of generality, we may assume that $b > a$. By our previous computation, the distance between z and w can be then expressed as

$$\mathrm{d}_{\mathbb{H}}(w, z) = \log \frac{\sin b}{\cos b + 1} - \log \frac{\sin a}{\cos a + 1}.$$

We would like express $\sin a$, $\cos a$, $\sin b$ and $\cos b$ in term of w, $\overline{w}$, z, $\overline{z}$. One way to to this is to compute x and r, then compute out the formula for the distance. However, this computation is very complicated. Hence, we would like to use some help from the Euclidean geometry. We consider the plane $\mathbb{C}$ endowed with Euclidean metric.

▼▼▼ **Recall some basic in Euclidean geometry**:

We recall two facts from the Euclidean geometry about inscribed angles and center angles. Let C be a circle with center O. Let p and q be two distinct points on C. They separate C into two arcs. We choose one point on each arc and denote them by A and B respectively. Without loss of generality, we assume that A and o are on the same side of the segment pq.

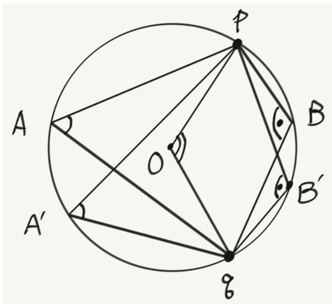

Figure 3.1.2: Inscribed angle and central angle

We consider only angles in $[0, \pi]$, and have the following two facts:

Proposition 3.1.2

$$\angle pAq + \angle pBq = \pi.$$

Proposition 3.1.3

$$2\angle pAq = \angle pOq.$$

Moreover if A' and B' are two points on the same arcs as A and B respectively, then we have

Proposition 3.1.4

$$\angle pA'q = \angle pAq,$$
$$\angle pB'q = \angle pBq.$$

▲▲▲

Now back to our question, notice that the center of C is on the real axis, hence $\overline{z}$ and $\overline{w}$ are on the same circle as z and w. Using above facts from Euclidean geometry, we have

$$\angle \overline{w} z w = a,$$
$$\angle \overline{z}\overline{w} z = \pi - b.$$

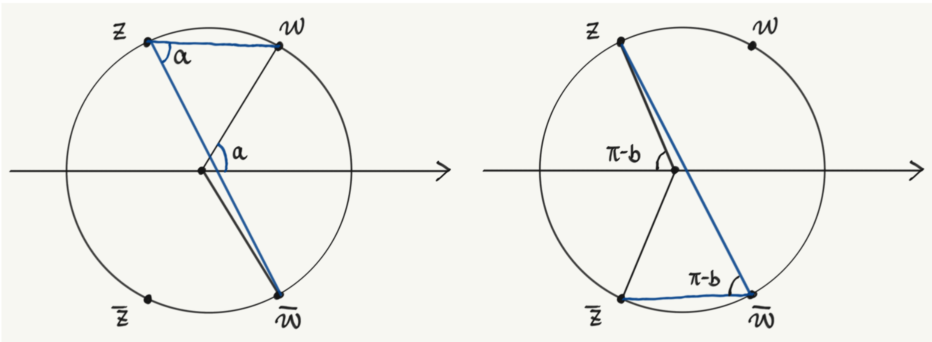

Figure 3.1.3: Equal angles

Using complex numbers, we have the following two relations:

$$\frac{w - z}{\overline{w} - z} = e^{ia} \frac{|w - z|}{|\overline{w} - z|},$$
$$\frac{\overline{z} - \overline{w}}{z - \overline{w}} = e^{i(\pi - b)} \frac{|\overline{z} - \overline{w}|}{|z - \overline{w}|}.$$

To simplify the notation, we use $z_1 = w - z$ and $z_2 = \overline{w} - z$. Notice that $z_1 + \overline{z_1} = z_2 + \overline{z_2}$. Then we have the following expressions:

$$e^{ia} = \frac{z_1 |z_2|}{z_2 |z_1|},$$
$$e^{ib} = -\frac{z_2 |z_1|}{\overline{z_1} |z_2|}.$$

Hence the two terms under the logarithm function in the distance formula can be expressed as:

$$\frac{\sin a}{\cos a + 1} = \frac{1}{i} \frac{e^{ia} - e^{-ia}}{e^{ia} + e^{-ia} + 2} = \frac{1}{i} \frac{z_1 |z_2| - z_2 |z_1|}{z_1 |z_2| + z_2 |z_1|},$$
$$\frac{\sin b}{\cos b + 1} = \frac{1}{i} \frac{e^{ib} - e^{-ib}}{e^{ib} + e^{-ib} + 2} = \frac{1}{i} \frac{\overline{z_1} |z_2| + z_2 |z_1|}{\overline{z_1} |z_2| - z_2 |z_1|}.$$

Hence we have

$$\mathrm{d}_{\mathbb{H}}(w, z) = \log \frac{|z_2|(z_1 + \overline{z_1}) + |z_1|(z_2 + \overline{z_2})}{|z_2|(z_1 + \overline{z_1}) - |z_1|(z_1 + \overline{z_1})} = \log \frac{|z_2| + |z_1|}{|z_2| - |z_1|}.$$

which is

$$\mathrm{d}_{\mathbb{H}}(w, z) = \log \frac{|\overline{w} - z| + |w - z|}{|\overline{w} - z| - |w - z|}.$$

When w and z are on the same vertical line, we have $\operatorname{Re} w = \operatorname{Re} z$. Hence we have

$$\begin{aligned} d_{\mathbb{H}}(w,z) &= \log \frac{|\overline{w}-z| + |w-z|}{|\overline{w}-z| - |w-z|} \\ &= \log \frac{(\operatorname{Im} w + \operatorname{Im} z) + |\operatorname{Im} w - \operatorname{Im} z|}{(\operatorname{Im} w + \operatorname{Im} z) - |\operatorname{Im} w - \operatorname{Im} z|} \\ &= \left| \log \frac{\operatorname{Im} w}{\operatorname{Im} z} \right|. \end{aligned}$$

Therefore the formula works in both cases. □

Remark 3.1.5.
For practical reasons, we can change it to a simpler form:

$$d_{\mathbb{H}}(w,z) = \log \left(\frac{2}{1 - \frac{|w-z|}{|\overline{w}-z|}} - 1 \right),$$

or consider one of the following formulas:

$$\begin{aligned} \cosh d_{\mathbb{H}}(w,z) &= 1 + \frac{|w-z|^2}{\operatorname{Im} w \operatorname{Im} z}, \\ \sinh \left(\frac{1}{2} d_{\mathbb{H}}(w,z) \right) &= \frac{|z-w|}{2(\operatorname{Im} w \operatorname{Im} z)^{1/2}}, \\ \cosh \left(\frac{1}{2} d_{\mathbb{H}}(w,z) \right) &= \frac{|z-\overline{w}|}{2(\operatorname{Im} w \operatorname{Im} z)^{1/2}}, \\ \tanh \left(\frac{1}{2} d_{\mathbb{H}}(w,z) \right) &= \frac{|z-w|}{|z-\overline{w}|}, \end{aligned}$$

This is because sometimes we do not care about the exact value of the distance, but rather the questions such as whether two distance are equal, or what is the derivative of the distance function as the points moving along certain path, etc. The above formula will simplify the computation for such questions. For example, to show

$$d_{\mathbb{H}}(w,z) = d_{\mathbb{H}}(w',z'),$$

we only need to verify the following relation:

$$\left| \frac{\overline{w}-z}{w-z} \right| = \left| \frac{\overline{w'}-z'}{w'-z'} \right|.$$

3.2 Distance function

Given pair of points w and z, we can fix w and move z along certain path, then the distance between w and z can be considered as a function on z. In this part, we would like to study this function.

3.2.1 Along horizontal lines H_y

Let $w = u + iv$ and $z = x + iy$. Let z move along the horizontal line H_y with imaginary coordinate y. We consider the formula

$$\cosh \mathrm{d}_{\mathbb{H}}(w, z) = 1 + \frac{|w - z|^2}{\operatorname{Im} w \operatorname{Im} z}.$$

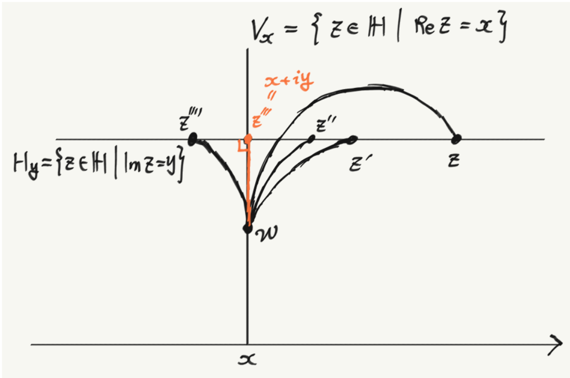

Figure 3.2.1: Distance to points in a horizontal line

Since we fix w and move z along the horizontal line, the real numbers u, v and y are constant, and the quantity $\cosh \mathrm{d}_{\mathbb{H}}(w, z)$ is a function $f(x)$ of x:

$$\begin{aligned} \cosh f(x) &= \cosh \mathrm{d}_{\mathbb{H}}(w, z) \\ &= 1 + \frac{|(u + iv) - (x + iy)|^2}{vy} \\ &= 1 + \frac{(u - x)^2 + (v - y)^2}{vy}. \end{aligned}$$

By the strict monotonicity and strict convexity of cosh function, and the strict convexity and the existence of the unique minimum of the quadratic polynomial function, we can already conclude that distance function has a unique minimum when $x = u$ and is strictly monotonic on both side of $x = u$.

Proposition 3.2.1
The function $f(x)$ has a unique minimum at $x = u$ and is strictly monotonic on its two sides.

Proof. We compute the first derivative of the length function.

$$\begin{aligned} \frac{\mathrm{d}(\cosh f(x))}{\mathrm{d}x} &= -\frac{2(u - x)}{vy}, \\ \frac{\mathrm{d}f(x)}{\mathrm{d}x} &= \frac{2(x - u)}{vy \sinh f(x)}. \end{aligned}$$

It is 0 if and only if $x = u$. □

Corollary 3.2.2
$f(x)$ will tend to infinite when x goes to $\pm\infty$.

Hence the distance between w and H_y is realized by the point $u + iy$. i.e.

$$\mathrm{d}_{\mathbb{H}}(w, H_y) = \mathrm{d}_{\mathbb{H}}(w, u + iy).$$

If w is not on H_y, the geodesic segment connecting w and $u + iy$ is on the vertical line V_u. Moreover $u + iy$ is the only point on H_y the geodesic connecting which to w is orthogonal to H_y.

3.2.2 Along vertical lines V_x

Let $w = u+iv$ and $z = x+iy$. Let z move along the vertical line V_x defined with real coordinate x.

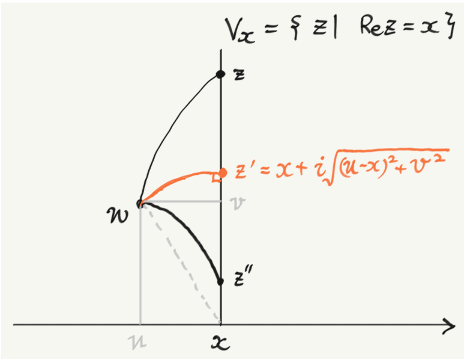

Figure 3.2.2: Distance to points in a vertical line

Since we fix w and move z along the horizontal line, the real numbers u, v and x are constant, and the quantity $\cosh \mathrm{d}_{\mathbb{H}}(w, z)$ is a function of y:

$$\begin{aligned}\cosh f(y) &= \cosh \mathrm{d}_{\mathbb{H}}(w, z)\\ &= 1 + \frac{|(u + iv) - (x + iy)|^2}{vy}\\ &= 1 + \frac{(u - x)^2 + (v - y)^2}{vy}.\end{aligned}$$

Proposition 3.2.3
The function $f(y)$ has a unique minimum.

Proof. The first derivative of $f(y)$ can be computed as follows:

$$\begin{aligned}\sinh f(y)\frac{\mathrm{d}f(y)}{\mathrm{d}y} &= \frac{1}{v}\frac{2(y-v)y-((u-x)^2+(v-y)^2)}{y^2} \\ &= \frac{1}{v}\frac{2y^2-2vy-(u-x)^2-v^2+2vy-y^2}{y^2} \\ &= \frac{1}{v}\frac{y^2-(u-x)^2-v^2}{y^2} \\ &= \frac{1}{v}\left(1-\frac{(u-x)^2+v^2}{y^2}\right)\end{aligned}$$

Since y is positive, we have the derivative of $D(y)$ is strictly monotonically increasing, and the minimum of $f(y)$ is realized by $y=\sqrt{((u-x)^2+v^2)}$. □

Corollary 3.2.4

The function $f(y)$ will tend to infinite when y goes to $+\infty$ and when y goes to 0.

Hence the distance between w and V_x is realized by the point $z'=x+i\sqrt{((u-x)^2+v^2)}$. i.e.

$$\mathrm{d}_{\mathbb{H}}(w,H)=\mathrm{d}_{\mathbb{H}}(w,z').$$

Notice that $|w-x|^2=((u-x)^2+v^2)$. Hence if w is not on V_x, then z' is the only point in V_x the geodesic connecting which to w is orthogonal to the vertical geodesic V_x.

3.2.3 Along circular geodesics $C(x,r)$

We consider the circular geodesic $C(x,r)$ centered at x with Euclidean radius r. Let $w=x+se^{i\xi}$ and $z=x+re^{i\theta}$.

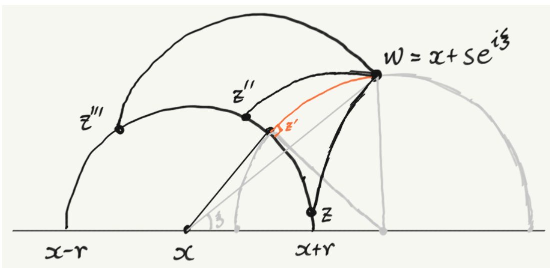

Figure 3.2.3: Distance to points in a circular geodesic

The distance between w and z is then a function $f(\theta)$ on θ.

$$\begin{aligned}\cosh f(\theta) &= \cosh \mathrm{d}_{\mathbb{H}}(w,z)\\ &= 1+\frac{|se^{i\xi}-re^{i\theta}|^2}{sr\sin\xi\sin\theta}\\ &= 1+\frac{(s\cos\xi-r\cos\theta)^2+(s\sin\xi-r\sin\theta)^2}{sr\sin\xi\sin\theta}\\ &= 1+\frac{s^2+r^2-2sr\cos(\theta-\xi)}{sr\sin\xi\sin\theta}.\end{aligned}$$

Proposition 3.2.5
The function $f(\theta)$ has a unique minimum.

Proof. We compute the first derivative of $D(\theta)$:

$$\begin{aligned}&\sinh f(\theta)\frac{\mathrm{d}f(\theta)}{\mathrm{d}\theta}\\ &= \frac{1}{sr\sin\xi}\frac{2sr\sin(\theta-\xi)\sin\theta-(s^2+r^2-2sr\cos(\theta-\xi))\cos\theta}{\sin^2\theta}\\ &= \frac{1}{sr\sin\xi}\frac{(s^2+r^2)\cos\theta-2sr\cos\xi}{\sin^2\theta}.\end{aligned}$$

Hence it is strictly monotonically increasing as θ goes from 0 to π. The unique minimum is given by

$$\cos\theta=\frac{2sr\cos\xi}{s^2+r^2}.$$

□

Corollary 3.2.6
The function $f(\theta)$ will tend to infinite when θ goes to 0 and when θ goes to π.

Let $z_{\min}$ denote the point on $C(x,r)$ realizing the distance between w and C. We consider the geodesic passing through $z_{\min}$ and w, and denote it by $C_{\min}(x_{\min},r_{\min})$. Using the Euclidean geometry, we have:

$$|z_{\min}-x_{\min}|^2=|w-x_{\min}|^2=r_{\min}^2.$$

Hence we have

$$\begin{aligned}x_{\min} &= \frac{|z_{\min}|^2-|w|^2}{2(\mathrm{Re}\,z_{\min}-\mathrm{Re}\,w)}\\ &= \frac{2x(r\cos\theta-s\cos\xi)+r^2-s^2}{2(r\cos\theta-s\cos\xi)}\\ &= x+\frac{r^2-s^2}{2(r\cos\theta-s\frac{(s^2+r^2)}{2sr}\cos\theta)}\\ &= x+\frac{r}{\cos\theta}.\end{aligned}$$

Hence the circle $C_{\min}(x_{\min}, r_{\min})$ intersects $C(x, r)$ orthogonally at $z_{\min}$ and $z_{\min}$ is the only point in $C(x, r) \cap \mathbb{H}$ having this property.

3.3 Applications

Corollary 3.3.1
Given any point w outside a geodesic γ, there is a unique geodesic passing w which intersects γ orthogonally. Moreover the geodesic segment between the intersection point and w realizes the distance between w and γ.

Definition 3.3.2
In all above cases, the unique point on the horizontal line H_y (resp. vertical geodesic V_x and circular geodesic $C(x, r)$) realizing the minimal distance from w is called the *(orthogonal) projection* of w to H_y (resp. V_x and $C(x, r)$).

Recall that the distance between two subsets of $\mathbb{H}$ is defined to be the infimum of the distance between points in them. Let K and K' be two subsets of $\mathbb{H}$. Assume that the distance $\mathrm{d}_{\mathbb{H}}(K, K')$ is realized by $\mathrm{d}_{\mathbb{H}}(z_K, z_{K'})$ with $z_K \in K$ and $z_{K'} \in K'$. Then we have

Proposition 3.3.3

$$\mathrm{d}_{\mathbb{H}}(K, K') = \mathrm{d}_{\mathbb{H}}(z_K, K') = \mathrm{d}_{\mathbb{H}}(K, z_{K'}).$$

Consider $K = \gamma$ and $K' = \eta$ are two geodesics, then we have

Corollary 3.3.4
If γ intersects η, then $\mathrm{d}_{\mathbb{H}}(\gamma, \eta) = 0$ is realized by the intersection point $z_\gamma = z_\eta$ is the intersection point.

Proof. This is due to the fact that the distance is always positive. Hence if there are pair of points from the two sets having distance 0, then it is the infimum. Moreover this is the only common point of γ and η, hence this pair realizing the distance 0 is unique. □

Corollary 3.3.5
The distance between two disjoint geodesics γ and η is realizable (i.e. z_γ and z_η exist), if and only if they do not share any end point in $\widehat{\mathbb{R}}$. Moreover the pair $(z_\gamma, z_\eta) \in \gamma \times \eta$, if exists, is unique.

Proof. We recall the following facts in Euclidean geometry:

(i) For any two circles C_1 and C_2, there is a unique circle C intersecting both of them orthogonally and the centers of the three circles are collinear if and only if C_1 and C_2 are disjoint.

(ii) For any line L and any circle C, there is a unique circle C' intersecting both of them orthogonally and the line L' passing through centers of the two circles is orthogonal to the line L if and only if L and C are disjoint.

Hence there is a unique geodesic orthogonal to both γ and η if and only if they are disjoint. We will use a deformation argument to show that the intersections of this geodesic with γ and η realize the distance between γ and η.

Using monotonicity of the distance function, we can find a segment I_γ on γ and a segment I_η on η such that the distance between I_γ and I_η is the same as that between γ and η. Now since $I_\gamma \times I_\eta$ is compact and the distance function is continuous, there exists a global minimal of distance function in $I_\gamma \times I_\eta$.

For any pair $(z, z') \in \gamma \times \eta$, we denote by $\sigma(z, z')$ the geodesic passing through z and z'. If any intersection angle between $\sigma(z, z')$ and γ or η is not $\pi/2$, there is a way to decrease the distance by moving either z or z'. On the other hand, if we take $\sigma(z, z')$ to be the common perpendicular geodesic of γ and η, then the intersection angles are both $\pi/2$. Hence locally deforming the intersection points z_γ and z_η will increase the distance. Hence z_γ and z_η realize a local minimum of the distance function, hence the global minimum. This means they realize the distance between γ and η. □

Remark 3.3.6.
When two geodesics share an end point, the distance between them is 0, but is not realizable. We will see this in more details later when we will talk about boundary at infinity of $\mathbb{H}$.

Let γ and η be two disjoint geodesics with no common end point.

Definition 3.3.7
The unique geodesic intersecting both γ and η orthogonally is called the *common perpendicular geodesic* of γ and η.

Remark 3.3.8.
Recall that in Euclidean geometry distance between two parallel lines can be realized by infinitely many pairs of points in them. The above discussion shows that in hyperbolic geometry this is no longer the case.

4 Circles in $\mathbb{H}$

4.1 Definitions

Another elementary geometric objects in $\mathbb{H}$ are circles. To see how they look like, we may try to compare them with the Euclidean circles. In fact, we will see that as subsets of $\mathbb{H}$, hyperbolic circles and Euclidean circles are the same. Let us first define what is a circle in $\mathbb{H}$ with respect to the hyperbolic metric.

Definition 4.1.1
A *circle* in $\mathbb{H}$ of radius R centered at w is the subset of $\mathbb{H}$ consisting of points with distance R to w:

$$C_{\mathbb{H}}(w,R) = \{z \in \mathbb{H} \mid \mathrm{d}_{\mathbb{H}}(z,w) = R\}.$$

A *open disk* in $\mathbb{H}$ of radius R centered at w is the subset of $\mathbb{H}$ consisting of points with distance R to w:

$$D_{\mathbb{H}}(w,R) = \{z \in \mathbb{H} \mid \mathrm{d}_{\mathbb{H}}(z,w) < R\}.$$

A *closed disk* in $\mathbb{H}$ of radius R centered at w is the subset of $\mathbb{H}$ consisting of points with distance R to w:

$$\overline{D}_{\mathbb{H}}(w,R) = \{z \in \mathbb{H} \mid \mathrm{d}_{\mathbb{H}}(z,w) \leq R\}.$$

4.2 Hyperbolic circles are also Euclidean circles

Let $w = u + iv$. Let $C = C_{\mathbb{H}}(w,R)$ be a circle of radius R centered at w. Using the convexity of the distance function, we can give a description of C by a sequence of propositions.

Proposition 4.2.1
The set $C_{\mathbb{H}}$ is non-empty for all w and all R.

Proof. Given any $w = u + iv \in \mathbb{H}$, we may consider the vertical line V_u. Then we may find two points $z^+ = u + ie^R v$ and $z^- = u + ie^{-R} v$ on V_u whose distance to w is R. Hence they are in C. □

We consider the intersection between C and horizontal lines in $\mathbb{H}$.

Proposition 4.2.2
The circle C is contained in the horizontal strip bounded by horizontal lines defined by $y = e^R v$ and $y = e^{-R} v$ respectively.

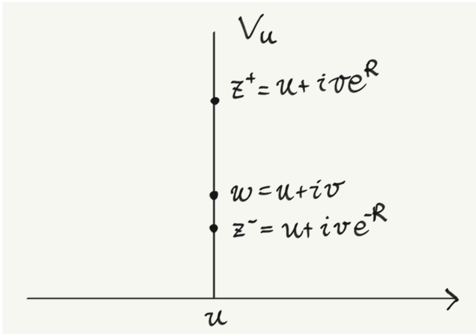

Figure 4.2.1: Highest and lowest points

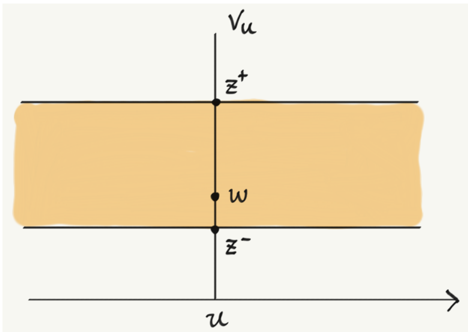

Figure 4.2.2: The strip containing the circle

Proof. This comes from our previous discussion on the convexity of the distance function for points moving along horizontal line. On any horizontal line H, the point realizing the minimal distance to w is the intersection $H \cap V_u$. By the distance formula for points on a same vertical line, we have the lemma. □

Proposition 4.2.3
For $e^{-R}v < y < e^R v$, there are exactly two points z_y^+ and z_y^- having distance R to w. Moreover we have

$$\frac{\mathrm{Re}\,(z_y^+ + z_y^-)}{2} = u,$$

i.e. they are symmetric with respect to $u + iy$ in H_y.

Without loss of generality, we may assume that $\mathrm{Re}\, z_y^- < \mathrm{Re}\, w < \mathrm{Re}\, z_y^+$.

Proof. This can be verified with the distance formula

$$\cosh \mathrm{d}_{\mathbb{H}}(w, z) = 1 + \frac{(u-x)^2 + (v-y)^2}{vy},$$

and the strict monotonicity of cosh on $\mathbb{R}_{\geq 0}$. □

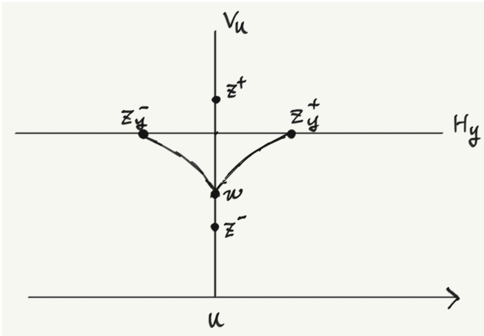

Figure 4.2.3: Intersection with a horizontal line in the strip

Since all points in $\mathbb{H}$ must be on some horizontal lines, by collecting $z^{\pm}$, and $z_y^{\pm}$'s for all $e^{-R}v < y < e^{R}v$, we have all points in C.

Our next step is to show that the circle in $\mathbb{H}$ is also a Euclidean circle. Since it is easier to compute hyperbolic distance from the Euclidean distance, we will show the following lemma, instead of a direct proof.

Proposition 4.2.4

All points on the Euclidean circle passing $z^{\pm}$ with center on V_u have hyperbolic distance R to w.

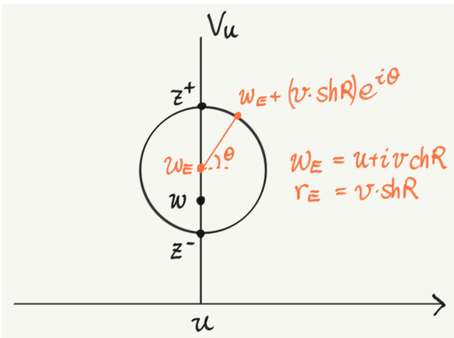

Figure 4.2.4: Parametrization of the Euclidean circle

Proof. Since the segment between z^- and z^+ is the diameter, we have the center of the Euclidean circle is $z_{\mathbb{E}} = u + iv\cosh R$ and the radius is $r = v\sinh R$. Hence the Euclidean circle can be described by

$$C(z_{\mathbb{E}}, r) = \{u + iv\cosh R + e^{i\theta}v\sinh R \in \mathbb{H} \mid \theta \in [0, 2\pi]\}.$$

Since there are two points on this circle having distance R to w, it is enough to show that the distance to w as a function of θ is constant, which is equivalent to show that

$$\frac{|z_{\mathbb{E}} + re^{i\theta} - w|}{|z_{\mathbb{E}} + re^{i\theta} - \overline{w}|} = \frac{|u + iv\cosh R + e^{i\theta}v\sinh R - w|}{|u + iv\cosh R + e^{i\theta}v\sinh R - \overline{w}|},$$

is constant. This can be see from the following computation:

$$\begin{aligned}
&\frac{|u+iv\cosh R+e^{i\theta}v\sinh R-w|}{|u+iv\cosh R+e^{i\theta}v\sinh R-\overline{w}|}\\
=&\frac{|u+iv\cosh R+e^{i\theta}v\sinh R-(u+iv)|}{|u+iv\cosh R+e^{i\theta}v\sinh R-(u-iv)|}\\
=&\frac{|iv(\cosh R-1)+e^{i\theta}v\sinh R|}{|iv(\cosh R+1)+e^{i\theta}v\sinh R|}\\
=&\frac{|2i\sinh^2\frac{R}{2}+2e^{i\theta}\sinh\frac{R}{2}\cosh\frac{R}{2}|}{|2i\cosh^2\frac{R}{2}+2e^{i\theta}\sinh\frac{R}{2}\cosh\frac{R}{2}|}\\
=&\left(\tanh\frac{R}{2}\right)\frac{|i\sinh\frac{R}{2}+e^{i\theta}\cosh\frac{R}{2}|}{|i\cosh\frac{R}{2}+e^{i\theta}\sinh\frac{R}{2}|}\\
=&\left(\tanh\frac{R}{2}\right)\frac{|-i\sinh\frac{R}{2}+e^{-i\theta}\cosh\frac{R}{2}|}{|i\cosh\frac{R}{2}+e^{i\theta}\sinh\frac{R}{2}|}\\
=&\left(\tanh\frac{R}{2}\right)\frac{|e^{i\theta}\sinh\frac{R}{2}+i\cosh\frac{R}{2}|}{|i\cosh\frac{R}{2}+e^{i\theta}\sinh\frac{R}{2}|}=\tanh\frac{R}{2}.
\end{aligned}$$

Hence the lemma. □

Since the intersection between the Euclidean circle and each horizontal line has also two points, we have the following proposition:

Proposition 4.2.5
As subsets of $\mathbb{H}$, we have $C_{\mathbb{H}}(w,R)=C(z_{\mathbb{E}},r)$ where

$$\begin{cases} z_{\mathbb{E}}=w\cosh R,\\ r=w\sinh R.\end{cases}$$

Corollary 4.2.6
The topology induced by Euclidean distance and that induced by hyperbolic metric is the same.

Corollary 4.2.7
As subsets of $\mathbb{H}$, we have $D_{\mathbb{H}}(w,R)=D(z_{\mathbb{E}},r)$ where

$$\begin{cases} z_{\mathbb{E}}=w\cosh R,\\ r=w\sinh R.\end{cases}$$

4.3 Circles and hyperbolic radius

Let C and C' be two hyperbolic circles of radius R an R' with a same center $z_{\mathbb{H}}=x+iy$. Without loss of generality, we assume that $R<R'$. Let w be a point on C, and γ be the geodesic passing $z_{\mathbb{H}}$ and w. Denote by w' the intersection $\gamma\cap C'$.

Proposition 4.3.1

$$d_{\mathbb{H}}(w, C') = d_{\mathbb{H}}(w, w').$$

Proof. Since the distance function is continuous and a circle is compact, the distance between w and C' is realized by some point $w'' \in C'$.

We prove by contradiction. Assume that $w'' \neq w'$. We consider the geodesic segment η connecting $z_{\mathbb{H}}$ and w''. Then $w \notin \eta$. We can build a path connecting $z_{\mathbb{H}}$ and w' by combining the geodesic segment $z_{\mathbb{H}}w$ and the geodesic segment ww''.

On one hand, we know that

$$d_{\mathbb{H}}(z_{\mathbb{H}}, w) + d_{\mathbb{H}}(w, w') = R' = d_{\mathbb{H}}(z_{\mathbb{H}}, w'').$$

On the other hand, by our assumption, we have $d_{\mathbb{H}}(w, w') > d_{\mathbb{H}}(w, w'')$. By the definition of distance or the triangular inequality, we know that

$$R = d_{\mathbb{H}}(z_{\mathbb{H}}, w'') \leq d_{\mathbb{H}}(z_{\mathbb{H}}, w) + d_{\mathbb{H}}(w, w'') < d_{\mathbb{H}}(z_{\mathbb{H}}, w) + d_{\mathbb{H}}(w, w') = R,$$

which is a contradiction. Hence the lemma. □

Corollary 4.3.2
Let C be a circle and $w \in \mathbb{H}$. The distance between w and C can be realized by a unique point on C which is the intersection between C and the geodesic determined by the center of C and w.

Proof. Let $z_{\mathbb{H}}$ be the hyperbolic center of C. Then w is on the circle with the same center of radius $d_{\mathbb{H}}(z_{\mathbb{H}}, w)$. By using the lemma, we have the corollary. □

Proposition 4.3.3
The geodesic segment $z_{\mathbb{H}}w$ is orthogonal to C at w.

Proof. Let C be a circle with hyperbolic center $z_{\mathbb{H}}$ and γ to be the unique geodesic tangent to C at w. It can be obtained in the following way. Let $z_{\mathbb{E}}$ be the Euclidean center of C. Then we consider the Euclidean line passing through $z_{\mathbb{E}}$ and w. Let x_γ denote its intersection with the real axis. Then the geodesic γ is the half circle centered at x_γ passing w.

Using Euclidean geometry, we know that γ only intersect the disk $D(z_{\mathbb{H}}, R)$ at the point w. Hence w realizing the minimal distance from $z_{\mathbb{H}}$ to γ. Hence the geodesic segment $z_{\mathbb{H}}w$ is orthogonal to γ, hence to C as well. □

We use the same notation as above. The parametrization of C is given by $iv \cosh R +$

$e^{i\theta}v\sinh R$. The length of C can be expressed as the following integral

$$\begin{aligned}
l_{\mathbb{H}}(C) &= \int_0^{2\pi} \frac{\sqrt{\dot{x}(\theta)^2+\dot{y}(\theta)^2}}{y(\theta)}\,\mathrm{d}\theta \\
&= \int_0^{2\pi} \frac{\sinh R}{\cosh R+\sinh R\sin\theta}\,\mathrm{d}\theta \\
&= \int_0^{\pi} \frac{\sinh R}{\cosh R+\sinh R\sin\theta}\,\mathrm{d}\theta + \int_{\pi}^{2\pi} \frac{\sinh R}{\cosh R+\sinh R\sin\theta}\,\mathrm{d}\theta \\
&= \int_{-\infty}^{\infty} \frac{\sinh R}{\cosh R+\sinh R\frac{2t}{t^2+1}}\frac{2}{t^2+1}\,\mathrm{d}t \\
&= \sinh R\int_{-\infty}^{\infty} \frac{2}{(1+t^2)\cosh R+2t\sinh R}\,\mathrm{d}t \\
&= \sinh R\int_{-\infty}^{\infty} \frac{2}{\cosh R+\cosh R(t+\tanh R)^2-\tanh^2 R\cosh R}\,\mathrm{d}t \\
&= \sinh R\int_{-\infty}^{\infty} \frac{2\cosh R}{(t\cosh R+\sinh R)^2+1}\,\mathrm{d}t \\
&= \sinh R\int_{-\infty}^{\infty} \frac{2}{u^2+1}\,\mathrm{d}u \\
&= 2\pi\sinh R.
\end{aligned}$$

Notice that given any point $z\in\mathbb{H}$ and any circle C with hyperbolic center z, any geodesic γ passing z will intersect C orthogonally. Given a disk of hyperbolic radius R, we consider it is foliated by circle of radius between 0 and R. Then the area of the disk of radius R can be computed as follows:

$$\begin{aligned}
A_{\mathbb{H}}(C) &= \int_0^R 2\pi\sinh t\,\mathrm{d}t \\
&= 2\pi(\cosh R-1).
\end{aligned}$$

Notice that when R goes to infinity, the perimeter of a disk and the area of a disk are comparable to each other.

Remark 4.3.4.
Later when we study the isometry group of $\mathbb{H}$, we will see that the rotations of $C_{\mathbb{H}}(w,R)$ with respect to w are isometries of $\mathbb{H}$. Hence the length is equidistribute on the circle with respect to the hyperbolic central angle, i.e. for any $\theta\in[0,2\pi]$, the arc on $C_{\mathbb{H}}(w,R)$ with hyperbolic central angle θ has length $\theta\sinh R$.

5 Horocycle and hypercycles

5.1 Circle revisit

Although circles in hyperbolic plane are the same as circles in Euclidean plane as subsets, there are still some difference between their geometric properties.

Let C be a circle in $\mathbb{H}$, which is symmetric with respect to a unique vertical line V_x with real coordinate x. Let $z_{\mathbb{H}} = x + iy_{\mathbb{H}}$ and $z_{\mathbb{E}} = x + iy_{\mathbb{E}}$ denote its hyperbolic and Euclidean center respectively. Let R and r denote its hyperbolic and Euclidean radius respectively.

With the hyperbolic data, we know that the diameter of C along V_C is the segment between $z^+ = x + ie^R y_{\mathbb{H}}$ and $z^- = x + ie^{-R} y_{\mathbb{H}}$. As the computation that we did above, we have the following relations:

$$y_{\mathbb{E}} = \frac{e^R y_{\mathbb{H}} + e^{-R} y_{\mathbb{H}}}{2} = y_{\mathbb{H}} \cosh R,$$
$$r = \frac{e^R y_{\mathbb{H}} - e^{-R} y_{\mathbb{H}}}{2} = y_{\mathbb{H}} \sinh R.$$

Using the relation between sinh and cosh functions, we have $y_{\mathbb{H}} = \sqrt{y_{\mathbb{E}}^2 - r^2}$. On the other hand, Hence the lower end of the vertical diameter has the coordinate

$$y_{\mathbb{H}} e^{-R} = y_{\mathbb{E}} - r.$$

We consider the ratio between $y_{\mathbb{H}}$ and $y_{\mathbb{H}} e^{-R}$:

$$e^R = \sqrt{\frac{y_{\mathbb{E}} + r}{y_{\mathbb{E}} - r}}$$

5.2 Horocycles

We would like to move the Euclidean circle along the vertical direction by Euclidean motions. In particular, we would like to decrease the y coordinate of points on the circle. From the above formula, we can see that both $y_{\mathbb{H}}$ and $y_{\mathbb{H}} e^{-R}$ go to 0, but $y_{\mathbb{H}} e^{-R}$ goes to 0 much faster than $y_{\mathbb{H}}$. When $y_{\mathbb{E}} = r$, we have $z_{\mathbb{H}} = x = z^-$ and $z^+ = x + 2ir$. Moreover, the hyperbolic radius is infinite.

Definition 5.2.1
Such a circle in hyperbolic space is called a *horocycle* centered at x passing through z^+.

In Euclidean geometry, the hyperbolic horocycle is a cycle tangent to the real axis and contained in $\mathbb{H}$.

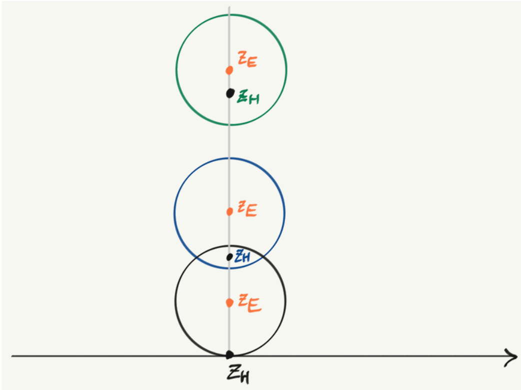

Figure 5.2.1: Drop a Euclidean ball

Remark 5.2.2.
Since the radius is infinite, we usually describe a horocycle by its center and the point it passes, as what we do for the Euclidean center. From the above point of view, a horocycle centered at x may be consider as a set of points which have the same "distance" to x, although the "distance" here is infinite.

Consider two concentric hyperbolic circles C and C' with hyperbolic center z of hyperbolic radius R and R' respectively. Given any geodesic passing through z, the part between the two circles has its length equals the difference between the two radius $|R - R'|$. Hence this quantity does not depend on the choice of the geodesic passing through z.

If we move C and C' simultaneously and keep them being concentric, then the above is still true. In the limit, we denote by H and H' the two horocycles. They share the same center x. Moreover all radius are issued from the center x on real axis. Although each radius is infinite, all geodesics issed from x, when restricted between H and H', will have the same length.

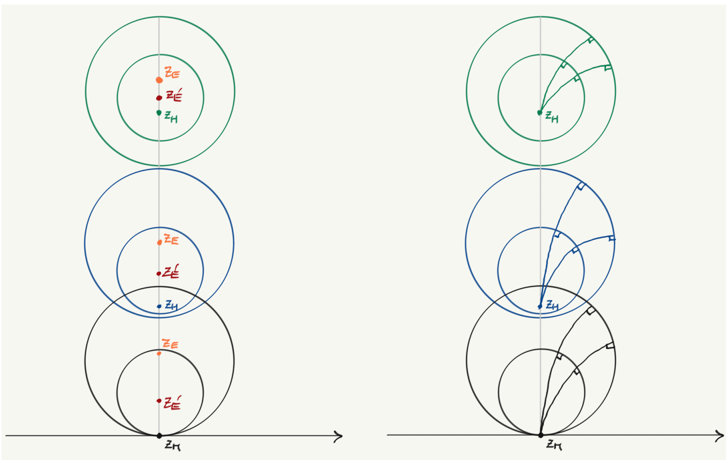

Figure 5.2.2: Drop 2 Euclidean ball at the same time

5.3 Hypercircles

One may ask after we get horocycle, what curve will we get if we move the Euclidean circle even lower. To start the discussion, we would like to check what happens when we move concentric circles downwards, especially the relation between their displacements.

Recall the relation between the hyperbolic data and Euclidean data that we computed in the beginning of this part:

$$y_{\mathbb{E}} = \frac{e^R y_{\mathbb{H}} + e^{-R} y_{\mathbb{H}}}{2} = y_{\mathbb{H}} \cosh R,$$
$$r = \frac{e^R y_{\mathbb{H}} - e^{-R} y_{\mathbb{H}}}{2} = y_{\mathbb{H}} \sinh R.$$

Hence we have $y_{\mathbb{H}} = \sqrt{y_{\mathbb{E}}^2 - r^2}$. Now we consider two concentric hyperbolic circles. Let us denote by $y_{\mathbb{E}}$ and $y'_{\mathbb{E}}$ their Euclidean center respectively, and by r and r' their Euclidean radius respectively. The concentric condition gives us a relation:

$$y_{\mathbb{E}}^2 - r^2 = y_{\mathbb{E}}'^2 - r'^2.$$

It holds while we are moving them to a lower position until they meet real axis. Since they are Euclidean circles and we consider the whole plane, there is still room to move them vertically lower. Both circles are no longer contained in $\mathbb{H}$ entirely. They intersect the real axis with two points. The above relation guarantees that the intersection points for different circles coincide.

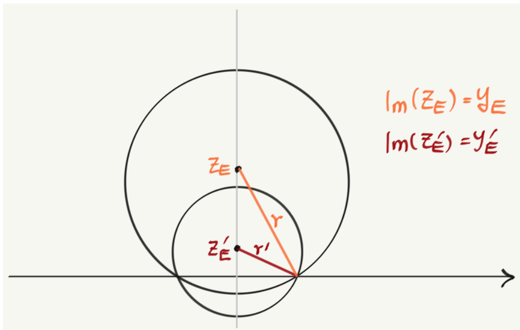

Figure 5.3.1: Drop under the horizon

Since there are two intersection points with the real axis, we have a unique complete geodesic γ associated to the two circles C and C'.

We first consider the circle C and γ. Let γ be the half circle given by $z = x + se^{i\theta}$. Let w be a point on $C \cap \mathbb{H}$. Its distance to γ is realized by a unique point z_0 on γ. We consider the geodesic γ_w passing w and z_0 given by $x_w + te^{i\xi}$. In order to compute the distance between w and z_0, we should have the angles $\alpha = \angle w x_w x$ and $\beta = \angle z_0 x_w x$. Our next step is to compute them.

Recall that the center of C is $z_{\mathbb{E}}$ and its Euclidean radius is r. We denote by y the imaginary art of $z_{\mathbb{E}}$. Using the relative position between C and γ, we have the relation $s^2 + y^2 = r^2$. Hence the Euclidean distance between $z_{\mathbb{E}}$ and w is also $s = \sqrt{r^2 - y^2}$. On the other hand, the geodesics γ and γ_w are orthogonal to each other, hence we have $(x_w - x)^2 = s^2 + t^2$. Hence the Euclidean distance between w and x_w is also $t = \sqrt{(x_w - x)^2 - s^2}$.

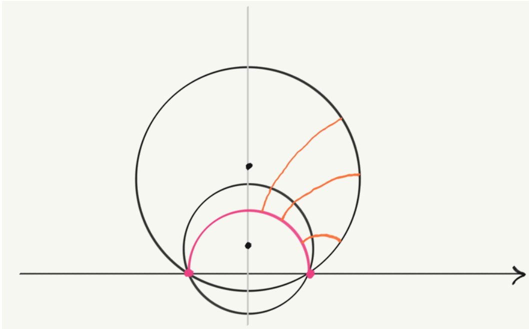

Figure 5.3.2: Geodesic associated to the two circles

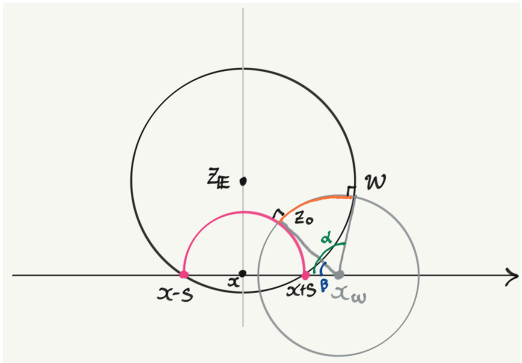

Figure 5.3.3: Distance from a point of C to γ

Now we consider the Euclidean triangle $\Delta z_\mathbb{E} x_w w$. Notice that the three side satisfying the Pythagorean equation:

$$|z_\mathbb{E} w|_\mathbb{E}^2 - r^2 - t^2 = y^2 + (x_w - x)^2 - (s^2 + y^2) - ((x_w - x)^2 - s^2) = 0$$

Hence the geodesic γ_w and C are orthogonal to each other.

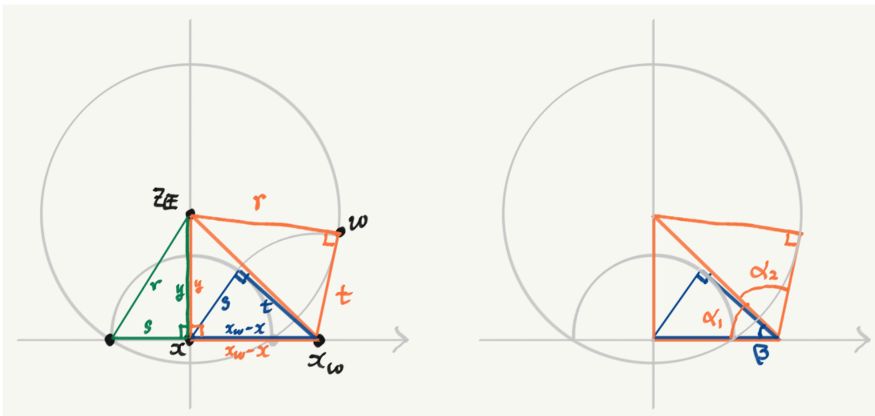

Figure 5.3.4: Quantities used in the computation

Now we would like to express $\sin\alpha$, $\cos\alpha$, $\sin\beta$ and $\cos\beta$ in term of y, s, r and x_w. The goal is to show that the distance between w and z_0 is independent of x_w. Hence all points on C have the same distance to γ. Using Euclidean geometry, we have the following formulas:

$$\begin{aligned}
\sin\beta &= \frac{s}{\sqrt{(x_w - x)^2}} \\
\cos\beta &= \frac{\sqrt{(x_w - x)^2 - s^2}}{\sqrt{(x_w - x)^2}} \\
\sin\alpha = \sin(\alpha_1 + \alpha_2) &= \frac{r(x_w - x)}{y^2 + (x_w - x)^2} + \frac{y\sqrt{(x_w - x)^2 - s^2}}{y^2 + (x_w - x)^2} \\
\cos\alpha = \cos(\alpha_1 + \alpha_2) &= \frac{\sqrt{(x_w - x)^2 - s^2}(x_w - x)}{y^2 + (x_w - x)^2} - \frac{yr}{y^2 + (x_w - x)^2}
\end{aligned}$$

Hence the distance can be given by

$$\begin{aligned}
\frac{\sin\alpha}{\cos\alpha + 1} &= \frac{r(x_w - x) + y\sqrt{(x_w - x)^2 - (r^2 - y^2)}}{((x_w - x) + \sqrt{(x_w - x)^2 - (r^2 - y^2)})(x_w - x) - y(r - y)} \\
\frac{\sin\beta}{\cos\beta + 1} &= \frac{\sqrt{r^2 - y^2}}{\sqrt{(x_w - x)^2 - (r^2 - y^2)} + x_w - x}
\end{aligned}$$

A direct computation shows

$$\frac{\sin\alpha}{\cos\alpha + 1}\frac{\cos\beta + 1}{\sin\beta} = \frac{y + r}{\sqrt{r^2 - y^2}},$$

hence is independent of x_w. Since x_w and w determine each other, this distance function is independent of the choice of w. Hence all points in $C \cap \mathbb{H}$ have the same distance to γ which is

$$\log\frac{y + r}{\sqrt{r^2 - y^2}} = \frac{1}{2}\log\frac{r + y}{r - y}.$$

Definition 5.3.1

A curve in $\mathbb{H}$ is called a *hypercircle* if it is obtained by the intersection between $\mathbb{H}$ and a Euclidean circle C which is not entirely contained in $\mathbb{H}$. The geodesic connecting the two points in the intersection between C and the real axis is called the *center* of this hypercircle.

Informally speaking, we may think that as the Euclidean circle moving towards the real axis and intersecting it, the hyperbolic center moves first from the interior of the disk bounded by the circle to the circle then moves out of it, in particular out of $\mathbb{H}$.

In fact this can be understood by considering the hyperbolic plane is part of the 2-dimensional projective space $\mathbb{RP}^2$. There is a natural duality between the hyperbolic space and the complement of its closure in $\mathbb{RP}^2$ coming from the $2+1$ Minkowski space. In particular, a point in the complement of $\mathbb{H}$ is dual to a geodesic in $\mathbb{H}$.

5.4 Horocycles and hypercircles with centers involving ∞

Instead of circle falling, we may consider circle raising. Let C be a circle with center on V_x. Let z^- denote the lower intersection between C and V_x. We would like to rise C while keeping the center on V_x and z^- constant. let us see how two centers and two radius changes during this process.

Recall the relations between the Euclidean data and the hyperbolic data:

$$\begin{aligned} y_{\mathbb{E}} &= \frac{e^R y_{\mathbb{H}} + e^{-R} y_{\mathbb{H}}}{2} = y_{\mathbb{H}} \cosh R, \\ r &= \frac{e^R y_{\mathbb{H}} - e^{-R} y_{\mathbb{H}}}{2} = y_{\mathbb{H}} \sinh R. \\ z^+ &= x + i e^R y_{\mathbb{H}} \\ z^- &= x + i e^{-R} y_{\mathbb{H}} \end{aligned}$$

where the hyperbolic center and the hyperbolic radius are $z_{\mathbb{H}} = x + iy$ and R respectively, and the Euclidean center and the Euclidean radius are $z_{\mathbb{E}} = x + iy_{\mathbb{E}}$ and r respectively. The points z^+ and z^- are the intersection points between the circle and the vertical line V_x.

Since we would like z^- to be constant. We have $e^{-R} y_{\mathbb{H}} = c$ for some constant $c > 0$. Hence we have

$$y_{\mathbb{H}} = c e^R,$$

which in turn induces

$$\begin{aligned} y_{\mathbb{E}} &= \frac{c(e^{2R} + 1)}{2}, \\ r &= \frac{c(e^{2R} - 1)}{2}. \end{aligned}$$

When we increase $y_{\mathbb{E}}$ to infinity, the Euclidean radius is getting bigger and bigger. In the limit, the circle converges to horizontal line. More precisely, if we consider an arc of a fixed length in the circle, during this process, this arc will converges to a segment on the horizontal line passing z^- in the Hausdorff distance for compact sets in $\mathbb{E}$.

At the same time, the hyperbolic center also goes to ∞. Hence by the similar argument, we can see that the horocycle centered at ∞ are horizontal lines. Moreover, the radius geodesics are vertical geodesics. Notice that the segments of vertical geodesics between two fixed horizontal lines have the same length, which we already saw this in the beginning when we computed the length of the vertical segments.

Now we consider hypercycles. We use a similar method to consider the vertical geodesic V_x as a limit of circular geodesics with one end fixed to be x while the other end tends to the positive infinity. More precisely, let γ be a circular geodesic with end point x and t with $x < t$. We would like to see how hypercycles changes when we move t to ∞ from the positive side.

Recall that as Euclidean circles, all hypercircles have angles with the geodesics at the end points. We would like to fix this intersection angle θ and see how the corresponding hypercircle changes when t goes to ∞. In fact the whole picture can be considered as applying rescaling the Euclidean plane with center at x. More precisely, since we fix the intersection angle between the hypercycle and the geodesic at x, the Euclidean center of the hypercircle will be on a Euclidean ray issued from x with angle θ to the positive direction of the real axis. As t goes to ∞, the Euclidean center of the hypercircle also moves to ∞ along the Euclidean ray.

We consider the convergence of subsets of the Euclidean plane, the hypercycle converges to the Euclidean ray issued from x with angle $(\pi/2) + \theta$. Hence the hypercircle centered at a vertical geodesic V_x is a ray issued from x. We have seen that the points on such a ray have a same distance to V_x in the beginning when we compute the length of a circular arc.

6 Boundary at infinity (Ideal boundary) of $\mathbb{H}$

We use $\widehat{\mathbb{R}}$ a lot in our previous discussion. In this part, we would like to show that $\widehat{\mathbb{R}}$ is indeed certain boundary of the hyperbolic space.

Consider a complete geodesic γ. For any point $z \in \gamma$, the complement of z consists of two connected components, each of which together with z is called a *geodesic ray* issued from z. Let γ^+ be a geodesic ray in γ. We will always consider the parametrization of γ^+ by the isometry between $[0, +\infty)$ and γ^+ (the *arc length parametrization*). Then for any pair of parameters t and t' in $[0, +\infty)$, we have

$$\mathrm{d}_{\mathbb{H}}(\gamma^+(t), \gamma^+(t')) = |t - t'|.$$

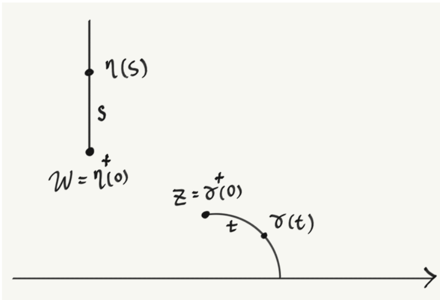

Figure 6.0.1: Geodesic rays with arc length parametrization

Definition 6.0.1

Two geodesic rays γ^+ and η^+ are said to be *equivalent to each other* if they stay in the bounded distance when going to infinite, i.e. there exists a constant $c > 0$, such that for any $t \in [0, +\infty)$, we have

$$\mathrm{d}_{\mathbb{H}}(\gamma^+(t), \eta^+(t)) \leq c.$$

Remark 6.0.2.

The key point is the existence of the constant c, instead of the actual value of c.

By triangular inequality of distance function, we can see that this definition is independent of the choice of the starting points of the geodesic rays on the complete geodesics.

Proposition 6.0.3

This define an equivalent relation among all geodesic rays.

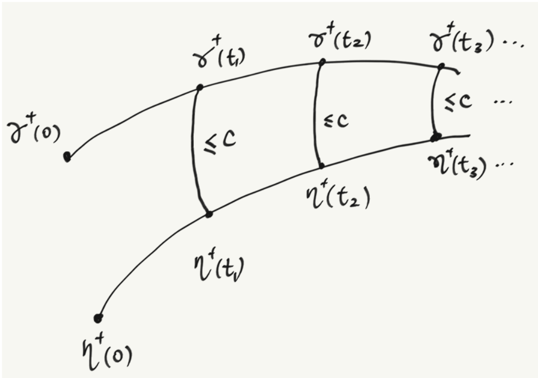

Figure 6.0.2: Equivalent geodesic rays

Proof. By triangular inequality of the distance function, we have this lemma. □

Definition 6.0.4
The *boundary at infinity (ideal boundary)* of $\mathbb{H}$ is defined to be the space of equivalence classes of all geodesic rays. We denote it by $\partial\mathbb{H}$. A point in $\partial\mathbb{H}$ is called an *ideal point*.

The following lemme shows that the ideal boundary of $\mathbb{H}$ can be identified with $\widehat{\mathbb{R}}$.

Proposition 6.0.5
Two rays are equivalent if and only if they end at a same point on the real axis or infinity.

Proof. We first consider x a point on the real axis, and show that all geodesics ending at x are equivalent to each other.

By our previous discussion on horocycles, all geodesics ending at x are the radius geodesics of horocycles centered at x. Their intersection with horocycles induce a natural identification between their parametrizations. Without loss of generality, we may assume that all rays ending at x are issued from points on a same horocycle. It is enough to show that all such rays are equivalent to the one in contained in the vertical geodesic V_x.

Let γ be any geodesic ending at x defined by $x + r + re^{i\theta}$. Let H be a horocycle centered at x given by $x + ir' + r'e^{i\theta}$. Therefore, the intersection $H \cap V_x$ is $z = x + 2ir'$. To see the intersection w between γ and H, we use the fact in Euclidean geometry that the segment between the intersections points of two circles is orthogonal to the segment between two centers

$$((x + r) - (x + ir')) \cdot (x + r + re^{i\theta} - x) = 0.$$

We compute the left hand side

$$\begin{aligned}
&((x + r) - (x + ir')) \cdot (x + r + re^{i\theta} - x) \\
=&(r - ir') \cdot (r + r\cos\theta + ir\sin\theta) \\
=&r^2(1 + \cos\theta) - rr'\sin\theta \\
=&r(r(1 + \cos\theta) - r'\sin\theta)
\end{aligned}$$

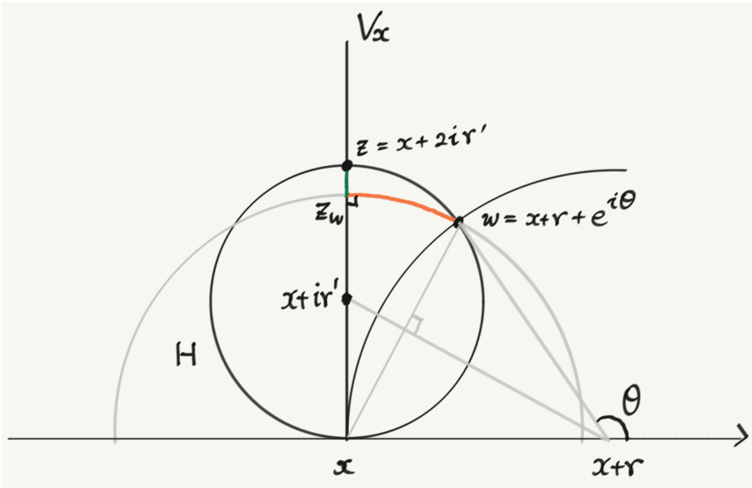

Figure 6.0.3: Equivalence rays ending at x

which yields

$$\tan\frac{\theta}{2} = \frac{r}{r'},$$
$$\cos\theta = -\frac{r^2 - r'^2}{r^2 + r'^2},$$
$$\sin\theta = \frac{2rr'}{r^2 + r'^2}.$$

Hence the intersection w can be written as

$$w = x + r + r\cos\theta + ir\sin\theta = x + r'\sin\theta + ir\sin\theta,$$

with $\tan(\theta/2) = r/r'$.

Let z_w denote the projection of w to V_x. It is given by

$$z_w = x + i\sqrt{(u-x)^2 + v^2} = x + i\sqrt{r'^2 + r^2}\sin\theta = x + i\frac{2rr'}{\sqrt{r^2 + r'^2}}.$$

By triangular inequality, we have

$$\mathrm{d}_{\mathbb{H}}(z, w) \le \mathrm{d}_{\mathbb{H}}(z, z_w) + \mathrm{d}_{\mathbb{H}}(z_w, w).$$

The distance between z_w and z is given by

$$\log\frac{\frac{2rr'}{\sqrt{r^2+r'^2}}}{2r'} = \log\frac{r}{\sqrt{r^2 + r'^2}},$$

which converges to 0 as r' goes to 0. The second term also converges to 0. This is because the angle $\angle z_w x w$ converges to 0. Recall the distance formula

$$\mathrm{d}_{\mathbb{H}}(z_w, w) = \log\frac{\cos\angle z_w x w}{\sin\angle z_w x w + 1},$$

which will converges to 0. Hence the two rays are equivalent, and so are all rays ending at x.

Now we consider all vertical geodesics. This case is easier than the previous case. Let V_x and $V_{x'}$ be two vertical geodesics, we consider their intersections with horizontal lines. Notice that for each such segment, the Euclidean distance is constant. Since they have to be renormalized by the inverse of the y coordinate, it will converges to 0. Hence the distance between them will converges to 0.

Now we have to show that if we have two rays γ and γ' ending at different points x and x' respectively, they are not equivalent. We consider the point x and all circular geodesics γ_r with Euclidean center x and Euclidean radius r. As r getting smaller and smaller, the ray ending x' will eventually stay on a different side of the geodesic γ_r from x. Recall the formula

$$\cosh \mathrm{d}_{\mathbb{H}}(w,z) = 1 + \frac{|w-z|^2}{\operatorname{Im} w \operatorname{Im} z}.$$

If x is on the real axis, then by the above formula, the points on γ_r will have their imaginary parts converges to 0 uniformly. Notice that the geodesic arc connecting corresponding points on γ and γ' will intersecting γ_r will eventually intersecting all γ_r, hence its length will be bigger than the distance $\mathrm{d}_{\mathbb{H}}(\gamma', \gamma_r)$ which tends to infinity. Hence these two rays are not equivalent. □

Corollary 6.0.6
The boundary at infinity $\partial\mathbb{H}$ is $\widehat{\mathbb{R}}$.

Next we would like to show that $\partial\mathbb{H}$ can be "attached to $\mathbb{H}$ nicely". We show the following two facts:

(i) The topology of $\mathbb{R}$ can be naturally extended to a topology on $\widehat{\mathbb{R}}$ which makes it homeomorphic to a circle.

(ii) We may extend the topology of $\mathbb{H}$ to a topology on $\overline{\mathbb{H}} = \mathbb{H} \cup \partial\mathbb{H}$, such that the induced topology on $\widehat{\mathbb{R}}$ is the same as the one as above.

The basis of the topology on $\widehat{\mathbb{R}}$ is given by the following two kinds of sets:

(i) open interval (x, x') with $x < x'$;

(ii) $(x,' \infty) \cup \{\infty\} \cup (\infty, x)$ with $x < x'$.

The sets of the first kind for a basis for the usual topology on $\mathbb{R}$. The sets of the second kind form a basis for the neighborhood of ∞.

To extend the topology of $\mathbb{H}$ to $\overline{\mathbb{H}}$, we consider half planes. The base of the topology of $\overline{\mathbb{H}}$ consists of two types subset:

(i) open disk in $\mathbb{H}$;

(ii) open half plane K determined by the geodesic with end points x and x' union with the open interval in the ideal boundary of K with end point x and x'.

7 Isometries of $\mathbb{H}$

7.1 Definition

To move objects in $\mathbb{H}$ without changing its geometric properties, we use isometris which are defined as follows:

Definition 7.1.1
A map f from $\mathbb{H}$ to itself is called an *isometry* if for any pair of points z and w, we have

$$\mathrm{d}_{\mathbb{H}}(f(z), f(w)) = \mathrm{d}_{\mathbb{H}}(z, w).$$

We denote by $\mathrm{Isom}(\mathbb{H})$ the set of all isometries. It is non-empty since the identity map is an isometry. There are some properties that we can get immediately from the definition.

Proposition 7.1.2
An isometry is continuous.

Using the continuity and triangular inequality, we have moreover:

Proposition 7.1.3
An isometry sends

- *geodesics to geodesics;*
- *circles to circles with the same radius.*

By the previous discussion, we then have:

Proposition 7.1.4
An isometry sends

- *horocycles to horocycles, preserving the distance among horocycles with the same center;*
- *hypercycles to hypercycles, preserving their distance to their center geodesic.*

Given any point w in $\mathbb{H}$, any point z in $\mathbb{H}$ is on a circle centered at w of the radius $\mathrm{d}_{\mathbb{H}}(w, z)$. Therefore, we have

Proposition 7.1.5
An isometry is bijective, and its inverse is also an isometry.

Proposition 7.1.6
The composition of two isometries is also an isometry.

Corollary 7.1.7
The set $\mathrm{Isom}(\mathbb{H})$ *has a group structure.*

We usually call it the *isometry group* of $\mathbb{H}$.

Let f be an isometry. Let K and K' be two non-empty subsets of $\mathbb{H}$, then an immediate consequence of the definition of an isometry is

Proposition 7.1.8

$$\mathrm{d}_{\mathbb{H}}(f(K), f(K')) = \mathrm{d}_{\mathbb{H}}(K, K').$$

In the following we will study some maps which can be easily verified to be isometries, then show that using them, we can get all isometries of $\mathbb{H}$.

7.2 First examples of isometries

We first study two elementary maps on $\mathbb{H}$:

$$T_t(z) = z + t$$
$$\phi_\lambda(z) = \lambda z$$

where $t \in \mathbb{R}$ and $\lambda > 0$ are both constant. The first is a Euclidean translation along the horizontal direction by a Euclidean distance t. The second is the Euclidean rescaling by a factor λ. The first thing that we would like to check is:

Proposition 7.2.1
These two maps act on $\mathbb{H}$.

Proof. It is enough to check the fact that they both preserve the property of the points in $\mathbb{H}$ that the imaginary part is positive. □

Then we would like to know if they preserve the distance.

Proposition 7.2.2
Both maps preserve distances between points on $\mathbb{H}$.

Proof. It is enough to check the following identity:

$$\left|\frac{z-w}{\overline{z}-w}\right| = \left|\frac{f(z)-f(w)}{f(\overline{z})-f(w)}\right|.$$

Notice that both the above map T_t and ϕ_λ are well defined on $\mathbb{C}$, hence the above identity is well defined for both of them. Let $f = T_t$, we have

$$\begin{aligned}&\left|\frac{T_t(z)-T_t(w)}{T_t(\overline{z})-T_t(w)}\right|\\ =&\left|\frac{(z+t)-(w+t)}{(\overline{z}+t)-(w+t)}\right|\\ =&\left|\frac{z-w}{\overline{z}-w}\right|.\end{aligned}$$

Let $f = \phi_\lambda$, we have

$$\begin{aligned}&\left|\frac{\phi_\lambda(z)-\phi_\lambda(w)}{\phi_\lambda(\overline{z})-\phi_\lambda(w)}\right|\\ =&\left|\frac{\lambda z-\lambda w)}{\lambda\overline{z}-\lambda w}\right|\\ =&\left|\frac{z-w}{\overline{z}-w}\right|.\end{aligned}$$

□

Another way to see these two maps are isometries on $\mathbb{H}$ is by comparing the path lengths before and after it acts. Let f be a map of class C^1 (map on $\mathbb{R}^2$) preserving $\mathbb{H}$, its derivative sends a tangent vector v based at z to a tangent vector $\mathrm{d}f(v)$ based at $f(z)$. For any pair of points z and w, the map f also sends paths connecting z and w to paths connecting $f(z)$ and $f(w)$. Hence if f-image of a regular path has the same length as that of the path itself, then it preserve the distance by the definition of the distance.

Let $\gamma(s) = (x(s), y(s))$ with $S \in [a, b]$ be a regular path connecting z and w two points in $\mathbb{H}$. Let $\eta(t) = f \circ \gamma(s)$ with $s \in [a, b]$ be the image path of γ under f. If $f = T_s$, we have $\eta(s) = (x(s) + t, y(s))$. Hence

$$|\dot{\eta}(s)|_{\mathbb{H}} = \frac{\sqrt{\dot{x}(s)^2 + \dot{y}(s)^2}}{y(s)} = |\gamma(s)|_{\mathbb{H}}.$$

If $f = \phi_\lambda$, we have $\eta(s) = (\lambda x(s), \lambda y(s))$. Hence

$$|\dot{\eta}(s)|_{\mathbb{H}} = \frac{\lambda\sqrt{\dot{x}(s)^2 + \dot{y}(s)^2}}{\lambda y(s)} = |\gamma(s)|_{\mathbb{H}}.$$

Hence in both cases we have $l_{\mathbb{H}}(\eta) = l_{\mathbb{H}}(\gamma)$.

Informally speaking, the map T_t translate objects along the horizontal direction without changing the size, and the hyperbolic metric has no change along the horizontal direction. Hence it is not hard to guess that T_t is an isometry. Let $\lambda > 1$, then the map ϕ_λ moves objects to a higher position. It rescales them into a big one, but the metric scales down by a same factor. These factors cancel out, hence it is also an isometry.

7.3 Reflections along geodesics in $\mathbb{H}$

We first consider the Euclidean reflection along the imaginary axis. The map is given by $\iota_0(z) = -\overline{z}$ which preserve $\mathbb{H}$.

Proposition 7.3.1
The map ι_0 is an isometry.

Proof. We verify this fact as follows:

$$\left|\frac{\iota_0(z) - \iota_0(w)}{\iota_0(\overline{z}) - \iota_0(w)}\right|$$
$$= \left|\frac{(-\overline{z}) - (-\overline{w})}{(-\overline{\overline{z}}) - (-\overline{w})}\right|$$
$$= \left|\frac{z - w}{\overline{z} - w}\right|.$$

□

Similar to the previous cases, the map ι_0 is also of class C^1. It sends (x, y) to $(-x, y)$. Hence if we consider a regular path $\gamma(s) = (x(s), y(s))$ and its image $\iota_0 \circ \gamma(s) = (-x(s), y(s))$. Then one may check that they have the same length.

One may wondering if we can also reflect $\mathbb{H}$ along a circular geodesic. We first consider the question: if there is a such an isometry, how should it looks like.

Let γ be the circular geodesic with center 0 passing i with the standard parametrization $\gamma(\theta) = e^{i\theta}$ with $\theta \in (0, \pi)$. We denote by ι_γ the reflection along γ that we are looking for. First since it is a reflection, we should have

Observation 7.3.2

$$\iota_\gamma(e^{i\theta}) = e^{i\theta}.$$

Secondly, notice that if a map preserve the distance between points, it also preserves the distance between subsets of $\mathbb{H}$. In particular, we have

Observation 7.3.3
For any point $z \in \mathbb{H}$, we have
$$\mathrm{d}_{\mathbb{H}}(z, \gamma) = \mathrm{d}_{\mathbb{H}}(\iota_\gamma(z), \gamma).$$
Moreover, the point z and its image $\iota_\gamma(z)$ have the same projection on γ.

Proof. The first part is because ι_γ is an isometry and preserves γ setwise. The second part is because ι_γ is an isometry and in fact fixes γ pointwise. □

Corollary 7.3.4
For any point $z \notin \gamma$, the geodesic passing z and $\iota_\gamma(z)$ is orthogonal to γ.

Proof. The geodesic η passing a point and its projection to a geodesic γ is orthogonal to γ. Since the projections of z and $\iota_\gamma(z)$ on γ are a same point, we have the corollary. □

Corollary 7.3.5
The map ι_γ preserves the imaginary axis.

Corollary 7.3.6
For any z, we have

$$\mathrm{d}_{\mathbb{H}}(z, V_0) = \mathrm{d}_{\mathbb{H}}(\iota_\gamma(z), V_0).$$

Remark 7.3.7.
In other words, the mao ι_γ preserves hypercycles centered at V_0.

For any point z, its distance to V_0 is realized by a circular arc with center at 0. Hence we have

Corollary 7.3.8
Any point z and its image $\iota_\gamma(z)$ are on a same Euclidean ray issued from 0. Moreover there exists a positive number λ, such that the projection of z to V_0 is $i\lambda$, while that of $\iota_\gamma(z)$ is $i\lambda^{-1}$.

The above necessary conditions for such an isometry existing suggest that this map should be the following one:

$$\iota_\gamma(z) = \frac{1}{\overline{z}}.$$

Proposition 7.3.9
The map ι_γ given by the above formula is an isometry.

Proof. We use the same idea as before.

$$\begin{aligned} & \left| \frac{\iota_\gamma(z) - \iota_\gamma(w)}{\iota_\gamma(\overline{z}) - \iota_\gamma(w)} \right| \\ = & \left| \frac{\overline{z}^{-1} - \overline{w}^{-1}}{\overline{\overline{z}}^{-1} - \overline{w}^{-1}} \right| \\ = & \left| \frac{z - w}{\overline{z} - w} \right|. \end{aligned}$$

□

The next important property of the reflection ι_γ is about the intersection angle between γ and other geodesics which intersects it. Let η be a geodesic different from γ which intersects γ at $z = e^{i\theta}$. Let x denote the end point of η between -1 and 1. Without loss of generality, we may assume that η is a circular geodesic. We will consider the case where η is a vertical geodesic as a limit case of the discussion.

Now we consider its image under the map ι_γ. Notice that it is well defined on $\mathbb{C} \setminus \{0\}$. We may formally define the image of 0 to be ∞. Hence the image of $\iota_\gamma(\eta)$ has one end point $1/x$. Without loss of generality, we may assume that $x > 0$. Let η^+ and γ^+ denote the geodesic rays issued from z to x and 1 respectively. Let α denote the angle between η^+ and γ^+, and let β denote the angle between $\iota_\gamma(\eta^+)$ and γ^+.

Proposition 7.3.10

$$\alpha = \beta.$$

Proof. Consider the vertical geodesic V passing z. We consider the geodesic ray V^+ issued from z and ending at $\cos\theta$. Let ξ_1 denote the angle between V^+ and η^+. We consider the circle containing η, and have the relation

$$\tan\frac{\xi_1}{2} = \frac{x - \cos\theta}{\sin\theta}$$

Let ξ_2 denote the angle between V^+ and $\iota_\gamma(\eta^+)$. We consider the circle containing η, and have the relation

$$\tan\frac{\xi_1}{2} = \frac{\frac{1}{x} - \cos\theta}{\sin\theta}$$

Notice that the angle between V^+ and γ^+ is θ. Hence to show $\alpha = \beta$, it is enough to show $\xi_1 + \xi_2 = 2\theta$. Since all angles are between 0 and π, it is enough to check

$$\tan\frac{\xi_1 + \xi_2}{2} = \tan\theta.$$

For this, we use the formula for computing tangent of a sum of two angles, and the left hand side is:

$$\begin{aligned}
&\tan\frac{\xi_1+\xi_2}{2} \\
=&\frac{\tan\frac{\xi_1}{2} + \tan\frac{\xi_2}{2}}{1 - \tan\frac{\xi_1}{2}\tan\frac{\xi_2}{2}} \\
=&\frac{\frac{x-\cos\theta}{\sin\theta} + \frac{\frac{1}{x}-\cos\theta}{\sin\theta}}{1 - \frac{x-\cos\theta}{\sin\theta}\frac{\frac{1}{x}-\cos\theta}{\sin\theta}} \\
=&\frac{(x + x^{-1})\sin\theta - 2\sin\theta\cos\theta}{\sin^2\theta - (1 - \cos\theta(x + x^{-1}) + \cos^2\theta)} \\
=&\tan\theta.
\end{aligned}$$

Notice that this also works for the case when η is a vertical geodesic. $\square$

Now we may consider to reflect $\mathbb{H}$ with respect to a different geodesic. We will consider to take conjugacy of ι_γ by T_t and ϕ_λ for some constant t and λ. We start by shows that by choosing t and λ, we can send γ to any circular geodesic in $\mathbb{H}$.

Proposition 7.3.11
Let η be a geodesic with center at x with Euclidean radius r. Then the isometry $f_\eta = T_x \circ \phi_r$ sends γ to η.

Proposition 7.3.12
The reflection of $\mathbb{H}$ along η can be expressed as follows:

$$\iota_\eta = f_\eta \circ \iota_\gamma \circ f_\eta^{-1}.$$

> If the map in the middle is identity map, then $f_\eta \circ f_\eta^{-1}$ is identity map. Informally speaking points are sent to some place then sent back following the same path but reversely, hence nothing is changed.
>
> If the map in the middle is a non identity map, then informally speaking points are sent from A to some place B, after moved in a non trivial way, sent back to A. The total effect of this process is the same as doing the middle part directly in the place A.

By the above proposition, the formula for the reflection at η (center at x with Euclidean radius r) is as follows:

$$\iota_\eta = \frac{r^2}{\overline{z} - x} + x = \frac{x\overline{z} - x^2 + r^2}{\overline{z} - x}.$$

If we fixes the left end point of η and move the right one to infinity, then we have $x - r = c$ a constant and

$$\lim_{x\to\infty} \frac{r + x}{x} = 2.$$

Hence for each z, in the limit we have $\iota_\eta(z) = -\overline{z} + 2c$. This is a reflection along the vertical geodesic V_c.

Proposition 7.3.13
For any geodesic η, the reflection $\iota_\eta : \mathbb{H} \to \mathbb{H}$ can be extend to a continuous map from $\overline{\mathbb{H}}$ to $\overline{\mathbb{H}}$.

7.4 Composition of two reflections

We consider η_1 and η_2 two geodesics intersecting each other at z. Let η_1^+ be a geodesic ray issued from z, and η_2^+ be a geodesic ray issued from z which is the one of the two geodesic rays from z that we meet first if we rotate η_1^+ in the positive direction. Let θ be the rotation angle from η_1^+ to η_2^+ along the positive direction. Let ι_1 be the reflection along η_1 and ι_2 be the reflection along η_2.

Proposition 7.4.1
The isometry $\iota_2 \circ \iota_1$ is a rotation of $\mathbb{H}$ by an angle 2θ to the positive direction around z.

Proof. A reflection along η will fixes all points on η, hence we have $\iota_2 \circ \iota_1(z) = z$, since $z \in \eta_1 \cap \eta_2$. Since $\iota_2 \circ \iota_1$ is an isometry, it preserves all circles centered at z.

Recall that a reflection along η preserves the intersection angles of geodesics with η. Hence all radius rays will be rotated to the positive direction by an angle 2θ. □

If η_1 and η_2 share one end point x, then the composition $\iota_2 \circ \iota_1$ preserves x and all horocycle centered at x. Let H be a horocycle. We consider the positive direction is induced by the positive orientation in $\mathbb{C}$. If the horocycle arc between η_1 and η_2 has the direct length t from η_1 to η_2, then $\iota_2 \circ \iota_1$ moves all points on H by a direct distance $2t$.

Example 7.4.2.
We will consider the case where $\{i\} = \eta_1 \cap \eta_2$. Let η_1 be the circular geodesic centered at 0, and η_2 be the geodesic centered at $x \in \widehat{\mathbb{R}}$. We first consider the case when $x \in \mathbb{R}$, then the case when $x = \infty$ can be considered as a limit case.

With the above assumption, the Euclidean radius of η_2 is $\sqrt{x^2+1}$. Let $\theta \in [0, \pi)$ denote the angle from η_1 to η_2. When $\theta \in [0, \pi/2]$, we have

$$\cos\theta = \frac{1}{\sqrt{x^2+1}},$$
$$\sin\theta = \frac{x}{\sqrt{x^2+1}}.$$

When $\theta \in [\pi/2, \pi)$, we have

$$\cos\theta = \frac{-1}{\sqrt{x^2+1}},$$
$$\sin\theta = \frac{-x}{\sqrt{x^2+1}}.$$

Hence we have the formula for the composition:

$$\begin{aligned}
&\rho_\theta(z) \\
=&(\iota_2 \circ \iota_1)(z) \\
=&\frac{x\overline{(\overline{z}^{-1})} - x^2 + r^2}{\overline{(\overline{z}^{-1})} - x} \\
=&\frac{xz^{-1} + 1}{z^{-1} - x} \\
=&\frac{z + x}{-xz + 1}
\end{aligned}$$

If we divided the denominator and numerator by $\sqrt{x^2+1}$ at the same time, we have

$$\rho_\theta(z) = \frac{\cos\theta z + \sin\theta}{-\sin\theta z + \cos\theta},$$

with $\theta \in [0, \pi)$. This is the isometry rotating $\mathbb{H}$ by an angle 2θ around i.

We consider the case where η_1 and η_2 are disjoint. Since for any η, geodesics orthogonal to η are preserved by the reflection ι_η, the common perpendicular geodesic σ of η_1 and η_2 is preserved by $f = \iota_2 \circ \iota_1$, and the end points of σ are also preserved. Since the isometry acts non trivially and isometrically on σ, its restriction σ is a translation ϕ. The translation distance is twice of $\mathrm{d}_{\mathbb{H}}(\eta_1, \eta_2)$. The direction is the same as from η_1 to η_2. Moreover, since it preserves σ, it also preserved all hypercycles centered at σ. The points on a hypercycle are moved in a way that their projections to σ are moved by the translation ϕ.

Example 7.4.3.
Let η_1 and η_2 be two circular geodesics centered 0 with Euclidean radius 1 and 2 respectively. Their common perpendicular geodesic is the imaginary axis. The distance between them is $\log 2$. The reflection along η_1 and η_2 are given by the following formulas:

$$\iota_1(z) = \frac{1}{\bar{z}}$$
$$\iota_2(z) = \frac{4}{\bar{z}}$$

Hence the composition $f(z) = (\iota_2 \circ \iota_1)(z) = 4z$ which is a rescaling. Notice that the only geodesic preserved by f is V_0 the imaginary axis. The action on it is translation following the direction from 0 to ∞ by distance $\log 4 = 2\log 2 = 2\mathrm{d}_{\mathbb{H}}(\eta_1, \eta_2)$.

Now we consider the last case when η_1 and η_2 share one end point x. Then the isometry $\iota_2 \circ \iota_1$ preserves x, hence all horocycles centered at x. Let H be a horocycle centered at x. Let the length of the arc of H between η_1 and η_2 is t. Then all points on H are moved by $\iota_2 \circ \iota_1$ following the direction from η_1 to η_2 for a distance $2t$ along H (the length of the arc in H between a point and its image is $2t$).

Example 7.4.4.
Let $\eta_1 = V_0$ and $\eta_2 = V_1$ be two vertical geodesics. They share a common end point ∞. Recall that horocycles centered at ∞ are horizontal lines. The reflections along η_1 and η_2 are given by

$$\iota_1(z) = -\bar{z}$$
$$\iota_2(z) = -\bar{z} + 2$$

Hence the composition $f(z) = (\iota_2 \circ \iota_1)(z) = z + 2$. We consider the horocycle H_1. The segment between η_1 and η_2 has length 1. Since we translate all point horizontally to the right by 2, hence for any point $z \in H_1$, the segment on H_1 between z and $f(z)$ has length 2.

For our convenience, we give names to the above three types of isometries obtained by composition of two reflections:

Definition 7.4.5 (i) When η_1 and η_2 intersects each other, we call $\iota_2 \circ \iota_1$ an isometry of *elliptic* type;

(ii) When η_1 and η_2 are disjoint, we call $\iota_2 \circ \iota_1$ an isometry of *hyperbolic* type;

(iii) When η_1 and η_2 are parallel, we call $\iota_2 \circ \iota_1$ an isometry of *parabolic* type.

Remark 7.4.6.
In particular, the isometries T_t's are of parabolic type, the isometries ϕ_λ's are of hyperbolic type, and the isometries ρ_θ's are of elliptic type.

7.5 All isometries can be expressed as compositions of reflections

In this part, we would like to show that using reflections is enough to produce all isometries of $\mathbb{H}$. More precisely, let η be a geodesic with end points x_1 and x_2 and $z_0 \in \eta$, we would like to show that an isometry f of $\mathbb{H}$ can be determined by the following data:

(i) the image $f(x_1)$ and $f(x_2)$,

(ii) the image $f(z_0)$,

(iii) the orientation of $f(\mathbb{H})$.

The following discussion including two steps. We will show first that there are two isometries sending (x_1, x_2, z_0) to $(f(x_1), f(x_2), f(z_0))$. Then the last condition on orientation tells the two candidates apart from each other.

For the first step, it is enough to show that there are exactly two isometries sending (x_1, x_2, z_0) to $(0, \infty, i)$. Recall that V_0 is the vertical geodesic with end point 0 and ∞. Let $z_0 = x + iy \in \eta$. We will follow the following three steps:

Step 1: Use the horizontal translation T_{-x} to send V_x to V_0. The intersection between V_0 and the image $T_{-x}(\eta)$ is iy.
Step 2: Use the rescaling of the Euclidean plane $\phi_{y^{-1}}$ to send iy to i.
Step 3: Let θ denote the angle between the ray issued from i ending at 0 and that ending at $x_1' = (\phi_{y^{-1}} \circ T_{-x})(x_1)$ along the positive direction. Use rotation $\rho_{-\frac{\theta}{2}}$ around i to send the ray ending at x_1' to the ray ending at 1.

By taking the composition, we have the isometry $f = \rho_{-\frac{\theta}{2}} \circ \phi_{y^{-1}} \circ T_{-x}$.

Proposition 7.5.1
For any geodesic η with end points x_1 and x_2, and any point $z_0 \in \eta$, there exist exactly two isometries f_η and $\widetilde{f}_\eta$, such that:

- *x_1 and x_2 are sent to 0 and ∞ respectively;*
- *z_0 is sent to i.*

Proof. Let f_1 and f_2 be two isometries satisfying the above two conditions. Then the composition $f = f_2 \circ f_1^{-1}$ is an isometry of $\mathbb{H}$ such that it fixes all points in V_0. This is because it preserves V_0, and fixes one point i. Hence its action on V_0 is either identity or reflection at i. Since the end points 0 and ∞ are also fixed, the restriction of the map f on γ is identity.

We consider i and $2i$ on γ. Then each point $w \in \mathbb{H}$ is on the hyperbolic circle centered at i of hyperbolic radius $d_{\mathbb{H}}(w, i)$ and on the hyperbolic circle centered at $2i$ of hyperbolic radius $d_{\mathbb{H}}(w, 2i)$. Since these circles are distinct Euclidean circles with Euclidean centers on V_0, the number of their intersection points is either 1 or 2. Moreover, if it is 1, then the two circles tangent to each other, hence the point $w \in V_0$. For any point w outside of V_0, the other intersection point is $-\overline{w}$.

Since an isometry is continuous, then $f(w) = w$ for all $w \in \mathbb{H}$ or $f(w) = -\overline{w}$ for all $w \in \mathbb{H}$. If $f = id$, then $f_1 = f_2$. If $f = \iota_0$ the reflection along V_0, then $f_2 = \iota_0 \circ f_1$. We denote by f_η the former and $\widetilde{f}_\eta$ the latter. □

Now back to the three conditions listed in the beginning. Let f be any isometry of $\mathbb{H}$. If we know $f(0)$, $f(\infty)$ and $f(i)$, then we have $f = f_\eta^{-1}$ or $f = \widetilde{f}_\eta^{-1}$ where η is the geodesics with end points $f(0)$ and $f(\infty)$ with orientation from 0 to ∞. The third condition about orientation will tell us which one of the two we should have.

Moreover, the above discussion also shows that all isometries can be obtained using compositions of ι_0, T_t's, ϕ_λ's, ρ_θ's. The isometries of the these three types can also be expressed as compositions of reflections. Hence

Corollary 7.5.2
All isometries can be expressed as a composition of some reflections of $\mathbb{H}$.

Corollary 7.5.3
All isometries preserve angles.

Proof. A reflection preserves angles. □

Remark 7.5.4.
For any one familiar with Lie groups, the algorithm describes the KAN decomposition of the Lie group $SL(2, \mathbb{R})$. The KAK decomposition of $SL(2, \mathbb{R})$ can bedescribed in a similar matter.

8 2×2 Matrices associated to isometries of $\mathbb{H}$

8.1 Matrices associated to reflections

From the discussion in last part, we can see that all isometries can be expressed as compositions of reflections. We recall that a reflection along a geodesic η of Euclidean radius r with center x can be expressed as:

$$\iota_\eta(z) = \frac{x\overline{z} + r^2 - x^2}{\overline{z} - x}.$$

To such a reflection, we may associate a 2×2 matrix A_η as follows

$$A_\eta = \begin{bmatrix} x & r^2 - x^2 \\ 1 & -x \end{bmatrix}.$$

When $\eta = V_c$ is a vertical geodesic, we then associate to ι_η the matrix

$$A_\eta = \begin{bmatrix} 1 & 2c \\ 0 & -1 \end{bmatrix}.$$

Let η_1 and η_2 be two geodesics. We denote their corresponding reflections by ι_1 and ι_2 respectively, and the matrices associated to these reflections by

$$A_1 = \begin{bmatrix} a_1 & b_1 \\ c_1 & d_1 \end{bmatrix}$$

and

$$A_2 = \begin{bmatrix} a_2 & b_2 \\ c_2 & d_2 \end{bmatrix}$$

respectively. We denote by $A = A_2 A_1$ their product and let

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} a_2a_1 + b_2c_1 & a_2b_1 + b_2d_1 \\ c_2a_1 + d_2c_1 & c_2b_1 + d_2d_1 \end{bmatrix}$$

Proposition 8.1.1
The isometry $f = \iota_2 \circ \iota_1$ is given by the following formula:

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

Proof. We compute the expression of f directly.

$$\begin{aligned} f(z) &= (\iota_2 \circ \iota_1)(z) \\ &= \iota_2(\iota_1(z)) \\ &= \iota_2\left(\frac{a_1\overline{z}+b_1}{c_1\overline{z}+d_1}\right) \\ &= \frac{a_2\overline{\frac{a_1\overline{z}+b_1}{c_1\overline{z}+d_1}}+b_2}{c_2\overline{\frac{a_1\overline{z}+b_1}{c_1\overline{z}+d_1}}+d_2} \\ &= \frac{a_2a_1z+a_2b_1+b_2c_1z+b_2d_1}{c_2a_1z+c_2b_1+d_2c_1+d_2d_1} \\ &= \frac{az+b}{cz+d}. \end{aligned}$$

□

More generally, let $\iota_1, ..., \iota_n$ be the reflections with respect to $\eta_1, ..., \eta_n$ respectively. Let $\psi : \mathbb{C} \to \mathbb{C}$ be the complex conjugate map. Let $A_1, ..., A_n$ denote the matrices associated to $\iota_1, ..., \iota_n$ respectively. Let

$$A = A_n \cdots A_1 = \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

Then we have

Proposition 8.1.2
The isometry $f = \iota_n \circ \cdots \circ \iota_1$ is given by the following formula:

$$f(z) = \psi^n\left(\frac{az+b}{cz+d}\right).$$

Proof. Let

$$A_\eta = \begin{bmatrix} s & t \\ p & q \end{bmatrix}$$

be the matrix associated to the reflection along η. Then the reflection ι_η can be written as the following composition:

$$\iota_\eta(z) = (\psi \circ h_\eta)(z) = \psi\left(\frac{sz+t}{pz+q}\right).$$

where $h_\eta(z) = (sz+t)/(pz+q)$. Moreover since all coefficient s, t, p and q are real numbers, we have ψ and h_η commute with each other.

Now we rewrite all reflections $\iota_1, ..., \iota_n$ as compositions $\psi \circ h_1, ..., \psi \circ h_n$. Then we have the following computation:

$$\begin{aligned} f &= \iota_n \circ \cdots \circ \iota_1 \\ &= (\psi \circ h_n) \circ \cdots \circ (\psi \circ h_1). \end{aligned}$$

Since ψ commutes with all $f_1, ..., f_n$, we can rewrite the above formula as follows

$$f = \psi^n \circ (f_n \circ \cdots \circ f_1).$$

Using a similar computation as in the proof of Proposition 8.1.1 and induction, we can show that

$$h_n \circ \cdots \circ h_1(z) = \frac{az+b}{cz+d}.$$

Hence we have the lemma.

□

Another observation is about the sign of the determinant of matrix A

Proposition 8.1.3
The determinant of A is non-zero, and it is positive if n is even and negative if n is odd.

Proof. Let A_η is the matrix associated to the reflection along η. If η is a circular geodesic, then the determinant of A_η is $-r^2$ where r is the Euclidean radius of the geodesic η which is non-zero. If η is a vertical geodesic, the determinant of A is -1. By the property of determinant, since A is the product of n matrices with strictly negative determinants, the determinant of A is non-zero and the sign depending on the number n: positive if n is even and negative if n is odd. □

Recall that the general linear group $\mathrm{GL}(2,\mathbb{R})$ consists of all 2×2 invertible matrices. The above lemma has the following immediate corollary.

Corollary 8.1.4
The matrix A belongs to $\mathrm{GL}(2,\mathbb{R})$.

8.2 Isometric action of $\mathrm{GL}(2,\mathbb{R})$ on $\mathbb{H}$

Reciprocally, given a matrix of $\mathrm{GL}(2,\mathbb{R})$, we may consider to define a map on $\mathbb{C} \setminus \mathbb{R}$ in a similar fashion. In the following part, we would like to show that in fact each such map when restricted to $\mathbb{H}$ is an isometry of $\mathbb{H}$.

More precisely, let

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

We define $h_A : \mathbb{C} \setminus \mathbb{R} \to \mathbb{C} \setminus \mathbb{R}$ by

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

Notice that there are two connected components of $\mathbb{C} \setminus \mathbb{R}$: the upper half plane $\mathbb{H}$ and the lower half plane $\mathbb{L}$.

Proposition 8.2.1
If $\det A > 0$, then h_A preserves $\mathbb{H}$ and $\mathbb{L}$ respectively. If $\det A < 0$, then h_A sends points in $\mathbb{H}$ (resp. $\mathbb{L}$) to $\mathbb{L}$ (reps. $\mathbb{H}$).

Proof. Notice that the difference between points in $\mathbb{H}$ and points in $\mathbb{L}$ is the sign of their imaginary part. Hence we consider the quantity $h_A(z) - \overline{h_A(z)}$ for each $z \in \mathbb{C} \setminus \mathbb{R}$.

$$\begin{aligned}
&h_A(z) - \overline{h_A(z)} \\
=&\frac{az+b}{cz+d} - \overline{\frac{az+b}{cz+d}} \\
=&\frac{(ac|z|^2 + adz + bc\bar{z} + bd) - (ac|z|^2 + ad\bar{z} + bcz + bd)}{|cz+d|^2} \\
=&\frac{(ad-bc)(z-\bar{z})}{|cz+d|^2} \\
=&\det A \frac{\operatorname{Im} z}{|cz+d|^2}
\end{aligned}$$

Hence if $\det A > 0$, then $h_A(z)$ and z have the same sign of the imaginary part. If $\det A < 0$, then $h_A(z)$ and z have different signs of the imaginary part. □

Notice that the complex conjugate exchanges $\mathbb{H}$ and $\mathbb{L}$. In order to define a map from $\mathbb{H}$ to itself when $\det A$ is negative, we take the composition of h_A with the complex conjugate map ψ. Moreover, since all coefficients in A are real, the map h_A and ψ commute with each other. We denote by $f_A = h_A$ if $\det A > 0$ and $f_A = h_A \circ \psi$ if $\det A < 0$. The above lemma shows that the restriction of f_A to $\mathbb{H}$ is well defined. We will denote this restriction also by f_A.

Proposition 8.2.2

For any $A \in \mathrm{GL}(2, \mathbb{R})$, the map f_A is an isometry of $\mathbb{H}$.

Proof. It is enough to check for any $A \in \mathrm{GL}(2, \mathbb{R})$, for any pair of points w and z in $\mathbb{H}$, we have:

$$\left| \frac{f_A(\overline{w}) - f_A(z)}{f_A(w) - f_A(z)} \right| = \left| \frac{\overline{w} - z}{w - z} \right|.$$

Since the complex conjugacy map ψ is complex linear, it is enough to check the following equality:

$$\left| \frac{h_A(\overline{w}) - h_A(z)}{h_A(w) - h_A(z)} \right| = \left| \frac{\overline{w} - z}{w - z} \right|.$$

The left hand side can be computed as follows:

$$\begin{aligned}
&\left| \frac{h_A(\overline{w}) - h_A(z)}{h_A(w) - h_A(z)} \right| \\
=&\left| \frac{a\overline{w}+b}{c\overline{w}+d} - \frac{az+b}{cz+d} \right| \Big/ \left| \frac{aw+b}{cw+d} - \frac{az+b}{cz+d} \right| \\
=&\left| \frac{(ac\overline{w}z + ad\overline{w} + bcz + bd) - (ac\overline{w}z + ad\overline{w} + bcz + bd)}{(acwz + adw + bcz + bd) - (acwz + adw + bcz + bd)} \right| \\
=&\left| \frac{(ad-bc)(\overline{w} - z)}{(ad-bc)(w-z)} \right| \\
=&\left| \frac{\overline{w} - z}{w - z} \right|.
\end{aligned}$$

Hence f_A is an isometry, for any $A \in \mathrm{GL}(2,\mathbb{R})$. □

Proposition 8.2.3
For any matrices A and B in $\mathrm{GL}(2,\mathbb{R})$, we have $f_{BA} = f_B \circ f_A$.

Proof. Since the complex conjugate map ψ commutes with h_A for any $A \in \mathrm{GL}(2,\mathbb{R})$, it is enough to check that for any A and B in $\mathrm{GL}(2,\mathbb{R})$, we have $h_{BA} = h_B \circ h_A$. This can be shown by a computation similar to the one in the proof of Proposition 8.1.1. □

Proposition 8.2.4
By sending $A \in \mathrm{GL}(2,\mathbb{R})$ to $f_A \in \mathrm{Isom}(\mathbb{H})$, we define a group homomorphism from $\mathrm{GL}(2,\mathbb{R})$ to $\mathrm{Isom}(\mathbb{H})$. It is surjective and the kernel is given by the subgroup of scalar matrices in $\mathrm{GL}(2,\mathbb{R})$.

Proof. Notice that the identity matrix $\mathrm{I}_2 \in \mathrm{GL}(2,\mathbb{R})$ is sent to identity map of $\mathbb{H}$ by definition. The previous lemma helps us check that the map sending A to f_A preserves the group structure. Hence this map is a group homomorphism.

The discussion in the previous section shows that any isometry f of $\mathbb{H}$ can be associated a matrix A_f. And the isometry f is the image of A_f under the group homomorphism. Hence this homomorphism is surjective.

We may check immediately that the subgroup of $\mathrm{GL}(2,\mathbb{R})$ consisting of scalar matrices is contained in the kernel.

$$f_{\lambda \mathrm{I}_2}(z) = \frac{\lambda z}{\lambda} = z.$$

To see it is all, we may consider the extension of the isometry f_A on $\mathbb{H}$ to a map on $\overline{\mathbb{H}}$. We will still denote it by f_A. In particular, we consider the f_A-image of 0 and ∞ and 1.

Let

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

Then $f_A(0) = b/d$, $f_A(\infty) = a/c$, and $f_A(1) = (a+b)/(c+d)$. Hence if A and B are not different by a scalar multiplication, then one of the ratio will be different. Hence one of the three geodesics connecting pairs of the three points 0, 1 and ∞ will sent to different places, which means that$f_A \neq f_B$ as maps in $\mathbb{H}$. Hence the kernel is the subgroup of scalar matrices in $\mathrm{GL}(2,\mathbb{R})$. □

We denote by $\mathrm{PGL}(2,\mathbb{R})$ the quotient group:

$$\mathrm{GL}(2,\mathbb{R})/\{\lambda \mathrm{I}_2 \mid \lambda \in \mathbb{R} \setminus \{0\}\}.$$

i.e. two matrices A and B are equivalent to each other if and only if there exists a non-zero real number λ such that $A = \lambda B$. This is not surprise, since we have

$$f_B(z) = f_{\lambda A}(z) = \frac{\lambda a z + \lambda b}{\lambda c z + \lambda d} = \frac{az+b}{cz+d} = f_A(z).$$

Another way to define the group $\mathrm{PGL}(2,\mathbb{R})$ is as follows:

$$\mathrm{PGL}(2,\mathbb{R}) := \{A \in \mathrm{GL}(2,\mathbb{R}) \mid |\det A| = 1\}/\{\pm \mathrm{I}_2\}.$$

It is called the *projective general linear group*.

Corollary 8.2.5
The isometry group $\mathrm{Isom}(\mathbb{H})$ *of* $\mathbb{H}$ *is isomorphic to* $\mathrm{PGL}(2, \mathbb{R})$.

Notice that the elements in $\mathrm{PGL}(2, \mathbb{R})$ can be classified into two types according to their determinant to be 1 or -1. In particular, those with determinant 1 form a subgroup of $\mathrm{PGL}(2, \mathbb{R})$, which is called the *projective special linear group*:

$$\mathrm{PSL}(2, \mathbb{R}) := \{A \in \mathrm{GL}(2, \mathbb{R}) \mid \det A = 1\}/\{\pm \mathrm{I}_2\}.$$

Recall that all matrices with determinant 1 for a subgroup of $\mathrm{GL}(2, \mathbb{R})$, which is called the *special linear group*:

$$\mathrm{SL}(2, \mathbb{R}) := \{A \in \mathrm{GL}(2, \mathbb{R}) \mid \det A = 1\}.$$

Hence we have $\mathrm{PSL}(2, \mathbb{R}) = \mathrm{SL}(2, \mathbb{R})/\{\pm \mathrm{I}_2\}$. By the discussion in the previous sections, we have

Corollary 8.2.6
The orientation preserving isometry group $\mathrm{Isom}^+(\mathbb{H})$ *of* $\mathbb{H}$ *is isomorphic to* $\mathrm{PSL}(2, \mathbb{R})$.

Definition 8.2.7
A *Möbius transformation* f on $\mathbb{H}$ is a map of the following form:

$$\begin{aligned} f : \mathbb{H} &\to \mathbb{H} \\ z &\mapsto \frac{az + b}{cz + d} \end{aligned}$$

where a, b, c and d are real numbers such that $ad - bc > 0$.

Alternatively, we may call the orientation preserving isometries of $\mathbb{H}$ the Möbius transformations on $\mathbb{H}$.

Remark 8.2.8.
Originally, a *Möbius transformation* is a conformal map from $\widehat{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$ (Riemann sphere) to itself. The general form of a Möbius transformation is as follows:

$$\begin{aligned} f : \widehat{\mathbb{C}} &\to \widehat{\mathbb{C}} \\ z &\mapsto \frac{az + b}{cz + d} \end{aligned}$$

where a, b, c and d are complex numbers such that $ad - bc \neq 0$.

The orientation preserving isometries considered here are precisely the restriction of those Möbius transformations on $\widehat{\mathbb{C}}$ preserving the upper half plane. Here we abuse the name and call them the Möbius transformations on $\mathbb{H}$.

Proposition 8.2.9

A Möbius transformation can be determined by the image of either one of the following three sets:

(i) two distinct points in $\mathbb{H}$;

(ii) one point in $\mathbb{H}$, and one point in $\partial\mathbb{H}$.

(iii) three distinct points in $\partial\mathbb{H}$.

Proof. Roughly speaking, the proof is given by a dimension counting.

Let $A \in \mathrm{SL}(2, \mathbb{R})$ be given by

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

The Möbius transformation f_A is determined by the coefficients a, b, c and d. Given any point $z \in \overline{\mathbb{H}}$ and its f_A-image $w \in \overline{\mathbb{H}}$, we have the following equality

$$\frac{az + b}{cz + d} = w,$$

which can be rewritten as

$$az + b = cwz + dw,$$

which is either a linear equation of a, b, c and d if $z \in \partial\mathbb{H}$, or two linear equations of a, b, c and d if $z \in \mathbb{H}$, by considering the real part and the imaginary part.

Notice that by choosing different z, we have non-equivalent linear equations for a, b, c and d. More precisely, let $f_A(z) = w$ and $f_A(z') = w'$. Then $z : 1 : zw : w = z' : 1 : z'w' : w'$ if and only if $z = z'$ and $w = w'$.

Since we have an extra equation $ad - bc = 1$, we only need three more equations, hence the lemma. □

Remark 8.2.10.

Another proof is given in the previous sections when we tried to show that all isometries can be obtained by composition of reflections. The rough idea is that if we know the image of two distinct points in $\mathbb{H}$, we know the images of points on the geodesic passing through this two points. Then each point out side of this geodesic can be determined by its distance to this geodesic and its projection to this geodesic. Hence the images of two points are enough to determine the Möbius transformation.

Similarly, the images of two distinct ideal points can determine the image of the geodesic connecting them up to a translation along the geodesic. By knowing image of one more point in $\overline{\mathbb{H}}$ different from these two ideal points, we know the image of its projection to this geodesic, hence the images of all points on this geodesic. Then we repeat the above discussion and obtained the image of all points in $\overline{\mathbb{H}}$, hence the Möbius transformation.

8.3 Classification of Möbius transformations

One advantage of considering matrices associated to isometries is that we can use the classification of matrices to describe the classification of Möbius transformations.

Previously, we have discussed all types of Möbius transformations which can be obtained by taking a compositions of two reflections. In fact, this is the classification of all Möbius transformation. We can see this using the Jordan form of a matrix. Recall that up to a multiplication with $-\mathrm{I}_2$, any matrix in $\mathrm{SL}(2,\mathbb{R})$ is conjugate to one of the following three forms:

$$\begin{bmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{bmatrix}, \quad \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}.$$

Hence any Möbius transformation can be conjugated to one which can be written as a composition of two reflections. Hence every non identity Möbius transformation is of one of the three types: hyperbolic, elliptic, parabolic.

Remark 8.3.1.
Since we consider 2×2 matrices, for any diagonal matrix, it can be conjugated to its inverse by

$$\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}.$$

Hence without loss of generality, we may always assume that $\lambda \geq 1$.

In the following we would like to see this in a different but related point of view. Notice that all Möbius transformations can be extended to a continuous map on $\overline{\mathbb{H}}$. By Brouwer fixed point theorem, each Möbius transformation has some fixed points in $\overline{\mathbb{H}}$.

Let

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \neq \pm \mathrm{I}_2$$

be a matrix in $\mathrm{SL}(2,\mathbb{R})$ and f_A be the associated Möbius transformation. We consider the following equation of z:

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

which can be rewritten as follows:

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

- If $c = 0$, then ∞ is one fixed point,
  - if $d - a \neq 0$, the point $z = -b/(d-a)$ is also a solution, hence we have two fixed points on $\partial\mathbb{H}$;
  - if $d - a = 0$, then there is no solution for the equation $(d-a)z + b = 0$, and ∞ is the only fixed point;
- If $c \neq 0$, then we have a quadratic equation with solutions z_1 and z_2 in $\mathbb{C}$,

- If $z_1 = z_2$, then this equation has only one real solution which is the fixed point of f_A in $\partial\mathbb{H}$;
- If z_1 and z_2 are different and both real, then they are the fixed points of f_A in $\partial\mathbb{H}$;
- If z_1 and z_2 are conjugate to each other, then there is only one contained in $\mathbb{H}$ which is the fixed point of f_A.

Remark 8.3.2.
To see ∞ is a fixed point, we may consider z_n a sequence of points in $\mathbb{H}$ tends to ∞, as n tend to infinity. If the value $f_A(z_n)$ also tends to ∞, as n tend to infinity, we say that ∞ is fixed by f_A.

By Proposition 8.2.9, we can conclude that there are three types of fixed point set of f_A:

- 2 distinct fixed points $\partial\mathbb{H}$;
- 1 fixed point in $\mathbb{H}$;
- 1 fixed point in $\partial\mathbb{H}$,

moreover, the fixed point(s) in each case is(are) the only fixed point(s) of f_A. Notice that f_A and its inverse have the same fixed points.

In particular, the type of set of fixed points induces a classification of f_A:

- In the first case, if x_1 and x_2 are two distinct fixed points of f_A in $\partial\mathbb{H}$, then f_A preserves the geodesic γ determined by x_1 and x_2, and it is hyperbolic;
- In the second case, if $z \in \mathbb{H}$ is the only fixed point of f_A, then f_A fixes all circles centered at z. It is a rotation of the hyperbolic plane, and hence elliptic;
- In the third case, if $x \in \partial\mathbb{H}$ is the only fixed point of f_A, then f_A must preserve all horocycles centered at x.

To see why this is true for the third case, we may assume that $x = \infty$. Then f_A sends horizontal lines to horizontal lines. Let $z = iy \in \mathbb{H}$. Assume that its image is $z' = x + iy'$. By compose T_{-x} to f_A, we move back z' to the imaginary axis, hence get an isometry preserving the imaginary axis. Hence it is a rescaling. Then f_A can be written as a composition $T_{-x} \circ \phi_\lambda$ where $\lambda = y'/y$. The expression of f_A is then $f_A(z) = \lambda z + x$. If $\lambda > 1$, then we consider the inverse of f_A which is $f_A^{-1}(z) = (z - x)/\lambda$. Notice that $f_A^{-n}(z)$ form a sequence in $\mathbb{H}$ which converges to a point in $\mathbb{R}$. Hence f_A^{-1} has a fixed point different from ∞ in $\mathbb{R}$. Hence f_A has two distinct fixed points which is a contradiction. Hence if f_A has one fixed point on $\partial\mathbb{H}$, it is parabolic.

We conclude as follows: a non identity Möbius transformation f_A is

- hyperbolic, if it has two distinct fixed points on $\partial\mathbb{H}$/it can be written as a decomposition of two reflections whose fixed geodesics are disjoint in $\overline{\mathbb{H}}$/we have $|\mathrm{tr}\, A| > 2$;
- elliptic, if it has one fixed points in $\mathbb{H}$/it can be written as a decomposition of two reflections whose fixed geodesics intersect each other/we have $|\mathrm{tr}\, A| < 2$;

- parabolic, if it has one fixed points on $\partial\mathbb{H}$/it can be written as a decomposition of two reflections whose fixed geodesics share an end point/we have $|\mathrm{tr}\, A| = 2$;

Definition 8.3.3

The geodesic preserved by a hyperbolic Möbius transformation f_A is called its *axis*.

8.4 Translation distance of a Möbius transformation

Let $A \in \mathrm{SL}(2, \mathbb{R})$ and f_A be the associated Möbius transformation.

Definition 8.4.1

The translation distance l_A of f_A is defined by the following quantity:

$$l_A = \inf\{\mathrm{d}_{\mathbb{H}}(z, f_A(z)) \mid z \in \mathbb{H}\}.$$

Proposition 8.4.2

If f_A is hyperbolic, then the translation distance is non zero and realized by the point on its axis.

Proof. Let η_A be the axis of f_A. Let w be any point in $\mathbb{H}$. We consider $f_A(w)$ and denote by z and $f_A(z)$ the projections of w and $f_A(w)$ to η_A respectively. Let γ_w denote the geodesic passing w and z which is orthogonal to η_A.

We have $\mathrm{d}_{\mathbb{H}}(w, f_A(w)) > \mathrm{d}_{\mathbb{H}}(z, f_A(z)) = \mathrm{d}_{\mathbb{H}}(\gamma_w, f_A(\gamma_w))$. On the other hand, since f_A is an isometry, all points on its axis η_A moved by a same directed distance along η_A. Hence the translation distance of f_A is realized by any point on η_A. □

If f_A is hyperbolic, the matrix A can be conjugated to the following one up to a multiplication with $-\mathrm{I}_2$:

$$\begin{bmatrix} \lambda_A & 0 \\ 0 & \lambda_A^{-1} \end{bmatrix},$$

where $\lambda_A > 1$. The corresponding Möbius transformation is ϕ_{λ_A}. Its axis is V_0. For any point $iy \in V_0$, its image is $i\lambda_A^2 y$. Hence we may verify the following relation:

$$\lambda_A = \exp\left(\frac{l_A}{2}\right).$$

Proposition 8.4.3

If f_A is elliptic, its translation distance is 0 and realized by its fixed point.

Proof. By definition, the translation distance l_A is always greater or equals to 0. Let z_A be its fixed point. It realizes the translation distance of l_A. Since it is the only fixed point of f_A, it is the only point realizes the translation distance of f_A. □

Proposition 8.4.4
If f_A is parabolic, the its translation distance is 0 and not realizable.

Proof. Let x be its fixed point on $\partial\mathbb{H}$. The map f_A sends a geodesic γ with one end point x to a geodesic denoted by $f_A(\gamma)$ which also has one end point x. Moreover, for any horocycle H centered at x, its intersection z with γ is sent to its intersection with $f_A(\gamma)$, denoted by $f(z)$.

In the previous part, when we discuss the boundary at infinity, we have shown that $d_{\mathbb{H}}(z, f(z))$ tends to 0 as z tends to x along γ. Hence the translation distance of f_A is 0. However, the only fixed point of f_A is x, hence there is no point in $\mathbb{H}$ realizing l_A. □

8.5 Triple transitivity of Möbius transformations on $\partial\mathbb{H}$

In this part, we would like to talk about the action of the group $\mathrm{Isom}(\mathbb{H})$ on $\partial\mathbb{H}$,

$$(\partial\mathbb{H})^{(2)} := \{(x, x') \in \partial\mathbb{H} \times \partial\mathbb{H} \mid x \neq x'\},$$

and

$$(\partial\mathbb{H})^{(3)} := \{(x, x', x'') \in \partial\mathbb{H} \times \partial\mathbb{H} \times \partial\mathbb{H} \mid x \neq x', \quad x' \neq x'', \quad x \neq x''\}.$$

In general, let G be a group and X be a set.

Definition 8.5.1
The group G admits an action on X if for each $g \in G$, there is a map $f_g : X \to X$, such that:

- for any x, we have $f_{id}(x) = x$;
- for any g and g', we have $f_{g'g} = f_{g'} \circ f_g$.

We assume that G acts on X.

Definition 8.5.2
An orbit of the G-action of x on X is the subset of X given by $\{f_g(x) \mid g \in G\}$.

Definition 8.5.3
The action of G on X is said to be *transitive* if for any pair of elements x and x' in X, there exists an element g of G, such that $f_g(x) = x'$.

Remark 8.5.4.
The action is transitive is equivalent to the fact that there is only one orbit.

Definition 8.5.5
The action of G on X is said to be *free*, if for any $x \in X$ and any $g \in G$, such that $f_g(x) = x$, then g is the identity element in G

Definition 8.5.6

For any element $x \in X$, the *stabilizer subgroup* G_x of G is defined to be the collection of elements g, such that $f_g(x) = x$.

Remark 8.5.7.

One can show that if a group G acts on X transitively, the stabilizer subgroups of points in X are isomorphic to each other. Moreover, the action is free if and only if the stabilizer subgroup of a point in X is trivial.

Our previous discussion suggests that there are well defined actions of $\mathrm{Isom}(\mathbb{H})$ on $\partial\mathbb{H}$, $(\partial\mathbb{H})^{(2)}$ and $(\partial\mathbb{H})^{(3)}$. All these actions are all transitive. Hence the stabilizer subgroups for different points isomorphic to each other.

Notice that the first two action are not free. The stabilizer subgroup for a point in $\partial\mathbb{H}$ is isomorphic to the upper triangular subgroup of $\mathrm{PGL}(2,\mathbb{R})$ (the stabilizer subgroup for ∞). The stabilizer subgroup of a point in $(\partial\mathbb{H})^{(2)}$ is isomorphic to the diagonal subgroup of $\mathrm{PGL}(2,\mathbb{R})$ (the stabilizer subgroup for $(0,\infty)$). The last action is free, i.e. the stabilizer subgroup of a point in $(\partial\mathbb{H})^{(3)}$ is trivial.

If we restrict ourselves to the subgroup $\mathrm{Isom}^+(\mathbb{H})$, then the action $\mathrm{Isom}^+(\mathbb{H})$ on $\partial\mathbb{H}$, $(\partial\mathbb{H})^{(2)}$ have similar property as $\mathrm{Isom}(\mathbb{H})$-action: being transitive; stabilizer subgroups isomorphic to upper triangular subgroup and diagonal subgroup of $\mathrm{PSL}(2,\mathbb{R})$ (not $\mathrm{PGL}(2,\mathbb{R})$) respectively.

The $\mathrm{Isom}^+(\mathbb{H})$-action on $(\partial\mathbb{H})^{(3)}$ is slightly different from the $\mathrm{Isom}(\mathbb{H})$-action. There are two orbits corresponding to the different cyclic order of a triple of distinct point on $\partial\mathbb{H}$, for example $(0,1,\infty)$ and $(0,\infty,1)$. We consider the orientation on $\partial\mathbb{H}$ following the orientation on the real axis, and denote by $(\partial\mathbb{H})^{(3)}_+$ the subset of $(\partial\mathbb{H})^{(3)}$ consisting of triples whose cyclic order coincide with the orientation on $\partial\mathbb{H}$. Then $\mathrm{Isom}^+(\mathbb{H})$-action on $(\partial\mathbb{H})^{(3)}_+$ is transitive and free.

8.6 Comments

From now on, it makes sense to talk about geometric objects in $\mathbb{H}$ without making its position precise. Alternatively, when we talk about a geometric object, we can always consider it to be at some position which is convenient for us to do computations. For example, later when we talk about disks or circles in the Poincaré disk model, it is always convenient to put the center at the origin; when we talk about triangles or more general polygons in $\mathbb{H}$, it is always convenient to put one side to be on the imaginary axis V_0.

9 Poincaré disk model $\mathbb{D}$

9.1 How to put the hyperbolic plane in the Euclidean plane in a conformal way

By our previous computation, the hyperbolic plane can be described using polar coordinates (R, θ) of the plane $\mathbb{R}^2$ with the following metric:

$$\mathrm{d}s_{\mathbb{H}}^2 = \mathrm{d}R^2 + \sinh^2 R \,\mathrm{d}\theta^2.$$

We would like to send this plane to the Euclidean plane, such that:

(i) The origin is sent to the origin;

(ii) A circle centered at the origin is sent to a circle centered at the origin keeping the central angle unchanged;

(iii) Radius are sent to radius.

(iv) The hyperbolic metric is written as $\mathrm{d}s_{\mathbb{H}}(R, \theta) = f(R)\,\mathrm{d}s_{\mathbb{E}}(r(R), \theta)$.

Recall that the Euclidean metric can be described as follows:

$$\mathrm{d}s_{\mathbb{E}}^2 = \mathrm{d}r^2 + r^2\,\mathrm{d}\theta^2.$$

If we try to send a hyperbolic circle of radius R to an Euclidean circle of radius R, then along the radius direction, the dilatation of the metric is 1 and along the circular direction, the dilatation factor is $R/\sinh R$.

On the other hand, if we try to send a hyperbolic circle of radius R to an Euclidean circle of the same length, then the dilatation of the metric along the circular direction is 1, but the dilatation of the metric along the radius direction is $\sinh R/R$.

Hence, we should consider a more complicated way to achieve our goal. Assume that the point $(R, \theta) \in \mathbb{H}$ is sent to $(r(R), \theta) \in \mathbb{E}$. Moreover, we assume that $r(R)$ is differentiable. Then we can rewrite the Euclidean metric as:

$$\mathrm{d}s_{\mathbb{E}}^2 = \left(\frac{\mathrm{d}r}{\mathrm{d}R}\right)^2 \mathrm{d}R^2 + r(R)\,\mathrm{d}\theta^2.$$

Then the hyperbolic metric is

$$\mathrm{d}s_{\mathbb{H}}^2 = f(R)^2 \left(\frac{\mathrm{d}r}{\mathrm{d}R}\right)^2 \mathrm{d}R^2 + f(R)^2 r^2(R)\,\mathrm{d}\theta^2 = \mathrm{d}R^2 + \sinh^2 R\,\mathrm{d}\theta^2.$$

Hence we have the following relation:

$$\frac{\mathrm{d}r}{r} = \frac{\mathrm{d}R}{\sinh R}.$$

By solving this ODE, we have

$$\log r(R) = \log c\frac{e^R - 1}{e^R + 1},$$

where $c > 0$ is a constant.

Without loss of generality, we may assume that $c = 1$. Then the point $(R, \theta) \in \mathbb{H}$ is sent to (r, θ) with

$$r(R) = \frac{e^R - 1}{e^R + 1}.$$

In particular, we can see that the range of r is $[0, 1)$. From this, we can express R in term of r:

$$R(r) = \log\frac{1 + r}{1 - r}.$$

Now we rewrite the hyperbolic metric with respect to the Euclidean parameters:

$$\begin{aligned} \mathrm{d}s_{\mathbb{H}}^2 &= \mathrm{d}R^2 + \sinh^2 R\,\mathrm{d}\theta^2 \\ &= \left(\frac{2}{1 - r^2}\right)^2 (\,\mathrm{d}r^2 + r^2\,\mathrm{d}\theta^2). \end{aligned}$$

9.2 Description of $\mathbb{D}$

The set of this model is given by the unit disk of $\mathbb{C}$:

$$\mathbb{D} = \{re^{i\theta} \in \mathbb{C} \mid 0 \leq r < 1, \quad 0 \leq \theta < 2\pi\}.$$

In this model, we use the polar coordinates of $\mathbb{C}$. The hyperbolic metric on $\mathbb{D}$ is given by

$$\mathrm{d}s_{\mathbb{D}} = \frac{2\sqrt{\mathrm{d}r^2 + r^2\,\mathrm{d}\theta^2}}{1 - r^2}.$$

Since both $\mathbb{H}$ and $\mathbb{D}$ are different models of a same space, one may wonder if there is a transformation map from one to the other. The answer is yes and the formula is given by

$$\begin{aligned} f_{\mathbb{D}} : \mathbb{H} &\to \mathbb{D}, \\ z &\mapsto \frac{z - i}{z + i}. \end{aligned}$$

Its inverse is given by

$$\begin{aligned} f_{\mathbb{D}}^{-1} : \mathbb{D} &\to \mathbb{H}, \\ z &\mapsto i\frac{1 + w}{1 - w}. \end{aligned}$$

The transformation $f_{\mathbb{D}}$ is called *Cayley transformation* and it is an isometry between the two models of the hyperbolic plane.

Remark 9.2.1.
Since we can compose an isometry of $\mathbb{D}$ and precompose an siometry of $\mathbb{H}$ to $f_{\mathbb{D}}$ to get new isometry sending $\mathbb{H}$ to $\mathbb{D}$, the isometry between these two models are far from being unique.

To see the connecting between $\mathbb{D}$ and $\mathbb{H}$, we consider the three dimensional Euclidean space. We denote by (x, y, t) the coordinate of a point. We consider the stereographic projection π of the unit sphere to P_{xy} the xy-plane from the point $(0, 0, 1)$.

If we consider HS_1 the half unit sphere defined by the condition $y > 0$, its image will be the half plane in P_{xy} given by $y > 0$ which we consider to be $\mathbb{H}$. If we consider HS_2 the half unit sphere defined by condition $t < 0$, the image is the unit disk in P_{xy} which we consider to be $\mathbb{D}$. Let ρ denote the rotation of R^3 fixing the line passing $(1, 0, 0)$ and $(-1, 0, 0)$, and it sends HS_1 to HS_2. Then the map from $\mathbb{H}$ to $\mathbb{D}$ can be written as the restriction of the composition $\pi \circ \rho \circ \pi^{-1}$ on $\mathbb{H}$.

Using the map $f_{\mathbb{D}}$, we may check that the complete geodesics in this model consists of two types:

(i) Diameters of $\mathbb{D}$;

(ii) Circular arcs on circles intersecting $\partial\mathbb{D}$ orthogonally.

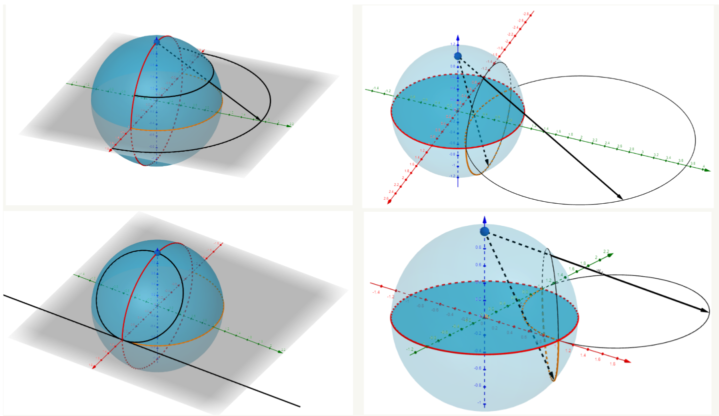

Figure 9.2.1: Geodesics in $\mathbb{H}$ and $\mathbb{D}$

For any points w and z in $\mathbb{D}$, the distance between them is given by

$$\mathrm{d}_{\mathbb{H}}(w, z) = \log \frac{|1 - \overline{w}z| + |w - z|}{|1 - \overline{w}z| - |w - z|}$$

Same as in $\mathbb{H}$, the circles in $\mathbb{D}$ are also Euclidean circles. Horocycles in $\mathbb{D}$ are circles tangent to a point on the unit circle. Given a geodesic γ in $\mathbb{D}$ with end point $e^{i\theta}$ and $e^{i\xi}$, all hypercircle centered at γ are given by the intersections between circles passing $e^{i\theta}$ and $e^{i\xi}$ with $\mathbb{D}$.

9.3 Isometries on $\mathbb{D}$

Since $f_{\mathbb{D}}$ is isometry from $\mathbb{H}$ to $\mathbb{D}$, given any isometry f on $\mathbb{H}$, the composition $f_{\mathbb{D}} \circ f \circ f_{\mathbb{D}}^{-1}$ is an isometry on $\mathbb{D}$. In this way we can get all isometry of $\mathbb{D}$. The isometries of $\mathbb{D}$ preserving the orientation are called the Möbius transformations on $\mathbb{D}$.

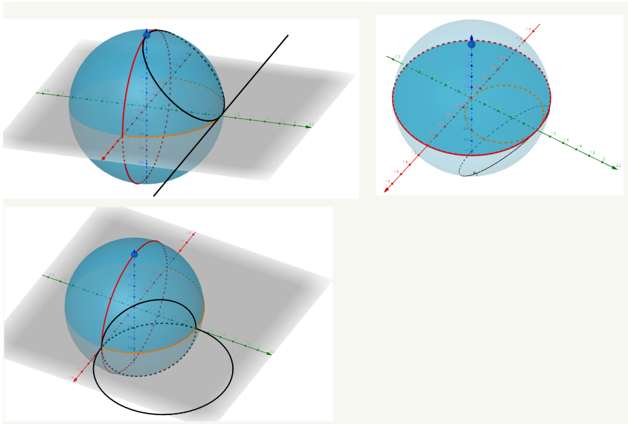

Figure 9.2.2: Horocycles in $\mathbb{H}$ and $\mathbb{D}$

Remark 9.3.1.
Similar to the Möbius transformation on $\mathbb{H}$, the Möbius transformations on $\mathbb{D}$ are the restrictions of the Möbius transformations of $\widehat{\mathbb{C}}$ preserving D on D.

Notice that the map $f_{\mathbb{D}}$ is also a Möbius transformation of $\widehat{\mathbb{C}}$. We may associated the matrix

$$A_{\mathbb{D}} = \begin{bmatrix} 1 & -i \\ 1 & i \end{bmatrix}$$

to $f_{\mathbb{D}}$. We consider the group $\mathrm{GL}(2,\mathbb{R})$ as a subgroup of $\mathrm{GL}(2,\mathbb{C})$, then we consider the group obtained by taking conjugacy of $\mathrm{SL}(2,\mathbb{R})$ by the $A_{\mathbb{D}}$. The resulting group is denoted by

$$\mathrm{U}(1,1) := \left\{ \begin{bmatrix} w & z \\ \overline{z} & \overline{w} \end{bmatrix} \in \mathrm{GL}(2,\mathbb{C}) \middle| |w|^2 - |z|^2 = 1 \right\}.$$

The orientation preserving isometry group of $\mathbb{D}$ is then isomorphic to

$$\mathrm{PU}(1,1) := \left\{ \begin{bmatrix} w & z \\ \overline{z} & \overline{w} \end{bmatrix} \in \mathrm{GL}(2,\mathbb{C}) \middle| |w|^2 - |z|^2 = 1 \right\} / \{\pm Id\}.$$

Remark 9.3.2.
The definition of $\mathrm{U}(1,1)$ is similar to the one for unitary group $U(2)$. We denote by J the following matrix:

$$\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

Then we may define $\mathrm{U}(1,1)$ as the following set:

$$\mathrm{U}(1,1) := \{M \in \mathrm{GL}(2,\mathbb{C}) \mid \overline{M}^t JM = J\},$$

where M^t is the transpose of M.

9.4 Compute lengths of circles and areas of disks

Up to applying an isometry, it is enough to consider circles with center at the origin. Let C be a circle with Euclidean radius r. Recall the following relations:

$$R = \log \frac{1+r}{1-r}.$$

and

$$r = \frac{e^R - 1}{e^R + 1} = \tanh \frac{R}{2}.$$

Then its length can be computed as follows:

$$\begin{aligned} l_{\mathbb{D}}(C) &= \int_0^{2\pi} \frac{2r\,\mathrm{d}\theta}{1-r^2} \\ &= \frac{4\pi r}{1-r^2} \\ &= \frac{4\pi \tanh \frac{R}{2}}{1-\tanh^2 \frac{R}{2}} \\ &= 2\pi \sinh R. \end{aligned}$$

The area of the disk bounded by C can be computed as before.

10 Cross ratios

In this part, we would like to introduce a useful tool in the study of hyperbolic geometry: cross ratios among ideal points.

10.1 See same directions from different reference points

Consider two geodesic rays starting from a point ending at two distinct ideal points. The angle between them may change as we move the starting point in $\mathbb{H}$. Hence if we try to give certain measurement on $\partial\mathbb{H}$ (for example visual angles from a point), it may not be canonic.

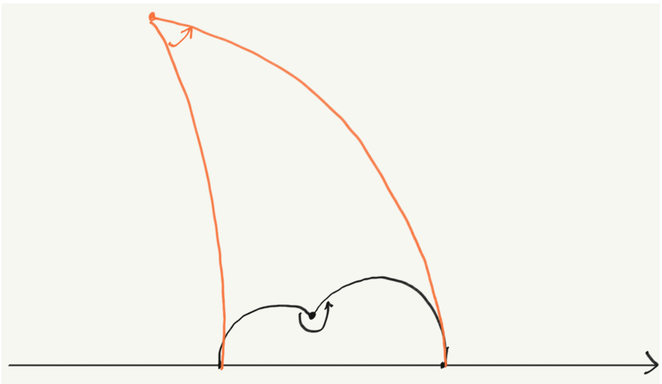

Figure 10.1.1: The visual angle depends on the reference point.

10.2 Quantities invariant under this change: Cross ratios

Given a reference point $z \in \mathbb{H}$ and 4 directions x_1, x_2, x_3 and x_4 on $\partial\mathbb{H}$, we denote by θ_1, θ_2, θ_3 and θ_4 the directed angles between the 4 directions with the negative vertical directions respectively.

We consider the following quantity associated to (x_1, x_2, x_3, x_4):

$$\mathbb{B}_z(x_1, x_2; x_3, x_4) := \frac{(\tan\frac{\theta_1}{2} - \tan\frac{\theta_4}{2})(\tan\frac{\theta_2}{2} - \tan\frac{\theta_3}{2})}{(\tan\frac{\theta_1}{2} - \tan\frac{\theta_3}{2})(\tan\frac{\theta_2}{2} - \tan\frac{\theta_4}{2})} = \frac{\sin\frac{\theta_1-\theta_4}{2}\sin\frac{\theta_2-\theta_3}{2}}{\sin\frac{\theta_1-\theta_3}{2}\sin\frac{\theta_2-\theta_4}{2}}.$$

Proposition 10.2.1

For any ordered quadruple points (x_1, x_2, x_3, x_4) on $\partial\mathbb{H}$, the quantity $\mathbb{B}_z(x_1, x_2; x_3, x_4)$ is independent of choice of z.

Proof. We define the following quantity

$$\mathbb{B}(x_1, x_2; x_3, x_4) := \frac{(x_1 - x_4)(x_2 - x_3)}{(x_1 - x_3)(x_2 - x_4)},$$

Using Euclidean geometry, we can conclude that for any points z and z' in $\mathbb{H}$, we have

$$\mathbb{B}_z(x_1, x_2; x_3, x_4) = \mathbb{B}(x_1, x_2; x_3, x_4) = \mathbb{B}_{z'}(x_1, x_2; x_3, x_4).$$

□

Definition 10.2.2
The quantity $\mathbb{B}(x_1, x_2; x_3, x_4)$ is called the *cross ratio* of the ordered quadruple (x_1, x_2, x_3, x_4).

Proposition 10.2.3
The function $\mathbb{B}$ *is invariant under Möbius transformations.*

Proof. We would like to show that for any 4 ordered ideal points (x_1, x_2, x_3, x_4), for any $z \in \mathbb{H}$, for any Möbius transformation f, we have

$$\mathbb{B}_z(x_1, x_2; x_3, x_4) = \mathbb{B}_{f(z)}(f(x_1), f(x_2); f(x_3), f(x_4)).$$

It is enough to show that for $z = i$, the quantity $\mathbb{B}_z(x_1, x_2; x_3, x_4)$ is invariant under T_t's, ϕ_λ's and ρ_θ's. The invariance under actions of T_t's and ϕ_λ is immediate, since they are Euclidean isometry and do not change angles.

Let $\theta \in [0, 2\pi]$. Then under the action ρ_θ, the four angles θ_i's becomes $\theta_i + 2\theta$'s. The following quantity stays constant under such a change:

$$\left(\sin\frac{\theta_1 - \theta_4}{2}\sin\frac{\theta_2 - \theta_3}{2}\right) \Big/ \left(\sin\frac{\theta_1 - \theta_3}{2}\sin\frac{\theta_2 - \theta_4}{2}\right)$$

□

The cross ratio has the following properties when we permute the points x_1, x_2, x_3 and x_4.

Proposition 10.2.4

$$\mathbb{B}(x_1, x_2; x_3, x_4) = \mathbb{B}(x_1, x_2; x_4, x_3)^{-1};$$
$$\mathbb{B}(x_1, x_2; x_3, x_4) = 1 - \mathbb{B}(x_1, x_3; x_2, x_4).$$

In fact there are more combinations $(4! = 24)$, if $\lambda = \mathbb{B}(x_1, x_2; x_3, x_4)$, then for any combination, the cross ratio is one of the following 6 values:

$$\lambda,\ 1 - \lambda,\ \frac{1}{\lambda},\ 1 - \frac{1}{\lambda},\ \frac{1}{1 - \lambda},\ 1 - \frac{1}{1 - \lambda}.$$

10.3 Geometric meaning of cross ratios

Since cross ratios are invariant under Möbius transformations, it is always convenient to renormalize the four ideal points to a standard position. Previously, we have shown that the action of $\mathrm{Isom}^+(\mathbb{H})$ is transitive and free on triple of distinct point on $\partial\mathbb{H}$. Hence without loss of generality, we may always assume that $(x_1, x_2, x_3) = (0, \infty, 1)$ or $(0, \infty, -1)$, depending on the cyclic order of (x_1, x_2, x_3).

If $(x_1, x_2, x_3) = (0, \infty, 1)$, we have

$$\mathbb{B}(0, \infty; 1, x_4) = x_4,$$

if $(x_1, x_2, x_3) = (0, \infty, -1)$, we have

$$\mathbb{B}(0, \infty; -1, x_4) = -x_4.$$

Proposition 10.3.1

Let η be the geodesic with end point x_1 and x_2, and let η' be the geodesic with end point x_3 and x_4, then η and η' intersect each other if and only if the cross ratio $\mathbb{B}(x_1, x_2; x_3, x_4)$ is negative.

To be more precise about the meaning of the cross ratio, we discuss case by case. We first consider the case $(x_1, x_2, x_3) = (0, \infty, 1)$. Let γ denote the geodesic with end points 1 and x_4. Hence if $\mathbb{B}(x_1, x_2; x_3, x_4)$ is positive, it equals to $\tanh^2(l/2)$ where l is the distance between γ and V_0. If $\mathbb{B}(x_1, x_2; x_3, x_4)$ is negative, it equals to $-\tan^2(\theta/2)$ where θ is the angle from V_0 to γ following the positive orientation.

We now consider the case $(x_1, x_2, x_3) = (0, \infty, -1)$. Let γ denote the geodesic with end points -1 and x_4. Hence if $\mathbb{B}(x_1, x_2; x_3, x_4)$ is positive, it equals to $\coth^2(l/2)$ where l is the distance between γ and V_0. If $\mathbb{B}(x_1, x_2; x_3, x_4)$ is negative, it equals to $-\cot^2(\theta/2)$ where θ is the angle from V_0 to γ following the positive orientation.

Proposition 10.3.2

Let γ be a geodesic with end point x_1 and x_2, and let η be a geodesic with end point x_3 and x_4, assume that (x_1, x_3, x_2) is in the positive cyclic order. Then the cross ratio $\mathbb{B}(x_1, x_2; x_3, x_4)$ is either $\tanh^2(l/2)$ where l is the distance between γ and V_0, if it is positive, or $-\tan^2(\theta/2)$ where θ is the angle from V_0 to γ following the positive orientation, if it is negative.

Another case that we would like to discuss here is related to hyperbolic Möbius transformation. Without loss of generality, we consider A to be the matrix

$$\begin{bmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{bmatrix},$$

where $\lambda > 1$. Then we have $f_A(z) = \lambda^2 z$. Its axis is V_0. Let x be any point on $\partial\mathbb{H}$ different from 0 and ∞. Then we have

Proposition 10.3.3

The cross ratio $\mathbb{B}(0, \infty; x, f_A(x)) = e^{l_A}$, where l_A is the translation distance of f_A.

Corollary 10.3.4

Let γ be a the axis of a hyperbolic Möbius transformation f with end points x_1 and x_2, such that the translation direction of f on γ is from x_1 to x_2. Let $x \in \partial\mathbb{H} \setminus \{x_1, x_2\}$. Then $\mathbb{B}(x_1, x_2; x, f(x)) = e^l$ where l is the translation distance of f.

Remark 10.3.5.

Cross ratios are widely used in the study of the (real or complex) projective geometry. In particular, we consider $\widehat{\mathbb{C}}$ as the complex 1-dimensional projective space, we may consider cross ratios for any distinct quadruple points on $\widehat{\mathbb{C}}$. In fact it is real if and only if the four points are on a same circle in $\widehat{\mathbb{C}}$. The cross ratio is invariant under the Möbius transformations of $\widehat{\mathbb{C}}$. Using the Cayley transformation, all results that we talked about above still holds using points in $\partial\mathbb{D}$.

11 Polygons

11.1 Triangles in $\mathbb{H}$

11.1.1 Definition

Definition 11.1.1
For any three distinct points z_1, z_2 and z_3 in $\mathbb{H}$, the *triangle* $\Delta(z_1, z_2, z_3)$ in $\mathbb{H}$ with vertices z_1, z_2 and z_3 is defined to be, as a subset of $\mathbb{H}$, the union of the three geodesics segments connecting pairs of z_1, z_2 and z_3. When z_1, z_2 and z_3 are on a same geodesic, we say that the triangle is *degenerate*.

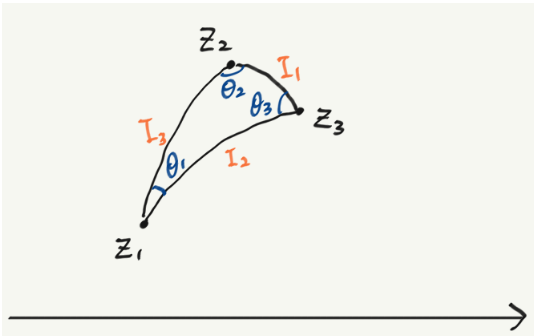

Figure 11.1.1: $\Delta(z_1, z_2, z_3)$

Definition 11.1.2
The three geodesic segments in a triangle are called its *sides*. The angles between pairs of sides with value in $[0, \pi]$ is called the *interior angles* of the triangle.

We denote by I_1, I_2 and I_3 the three sides opposite to the three vertices z_1, z_2 and z_3 respectively, and by l_1, l_2 and l_3 their lengths respectively. We denote by θ_1, θ_2 and θ_3 the three interior angles associated to z_1, z_2 and z_3 respectively.

Definition 11.1.3
By allowing vertices to be on $\partial\mathbb{H}$, we get *triangles with ideal points*. In particular, a triangle with three ideal points is called an *ideal triangle*.

When the vertex is idea, as a convention, we say that the interior angles is 0 at this vertex.

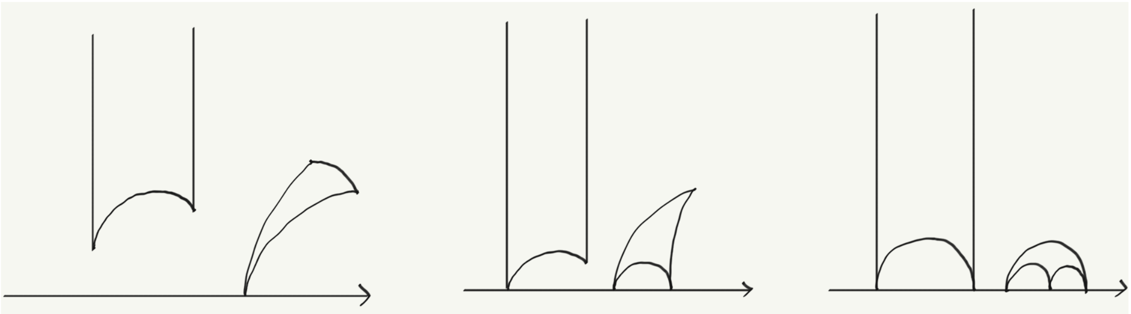

Figure 11.1.2: Triangles with ideal points

11.1.2 Determine a triangle by its interior angles

When there is an ideal vertex in a triangle, its adjacent sides will have infinite lengths which are not easy to compare with each other. On the other hand, it is easy to describe the angles of triangles. We would like to show in this part that the interior angles can be used to determine a triangle. In the other words, we would like prove the following proposition:

Proposition 11.1.4
Two hyperbolic triangles are isometric to each other if and only if their interior angles are the same.

Remark 11.1.5.
This proposition shows us that unlike in Euclidean geometry, we do not have similarity of triangles in hyperbolic geometry.

Similar to the discussion on distances, we would like to show some strict monotonicity of angles when we move points along geodesics. Let us be more precise.

We first consider the case when $\Delta(z_1, z_2, z_3)$ has no ideal point. We denote by γ_1, γ_2 and γ_3 the three geodesics containing I_1, I_2 and I_3 respectively. By applying an isometry, we may assume that $z_1 = i$ and $\gamma_2 = V_0$, such that z_2 has positive real part. We will try to construct the triangle $\Delta(z_1, z_2, z_3)$ by determining the position of γ_1. Recall that we have the following fact.

Proposition 11.1.6
Given a geodesic γ, a any geodesic η intersecting γ can be determined by its intersecting point and its intersecting angle.

Proof. To each such data, we have a unique point $z \in \mathbb{H}$ and a unique direction $\overrightarrow{v}$ at z. It determines a unique Euclidean straight line L. Now we consider all circles tangent to this line with center on $\mathbb{R}$. To do so, we may consider the Euclidean line L' orthogonal to L at z. Then its intersection with the real axis is the center that we are looking for. Hence we obtain a geodesic in $\mathbb{H}$. Notice that this construction only produce one unique geodesic in $\mathbb{H}$, hence the proposition. □

Remark 11.1.7.
This is essentially to say that for any tangent vector $(z, \overrightarrow{v})$ of $\mathbb{H}$, there is a unique geodesic passing z following the direction of $\overrightarrow{v}$.

We move the point $z_1 = i$ along V_0 to the positive direction of V_0. We denote the resulting point by z. Let γ_z denote the geodesic intersecting V_0 with angle θ_3. Let θ_z denote the intersecting angle between γ_z and γ_3.

Proposition 11.1.8
The angle θ_z as a function on $z = iy$ is strictly monotonically decreasing to 0 as y increasing.

Proof. We may try to compute the formula for θ_z using Euclidean geometry. To do so, we first get the position of the geodesic γ_z. Notice that it intersects V_0 with angle θ_3. Let $z = iy$. Then the center of γ_z is $x_z = -y/\tan\theta_3$ and the radius is $r_z = y/\sin\theta_3$. On the other hand, the geodesic γ_3 has center at $x_3 = 1/|\tan\theta_1|$ and radius $r_3 = 1/\sin\theta_1$.

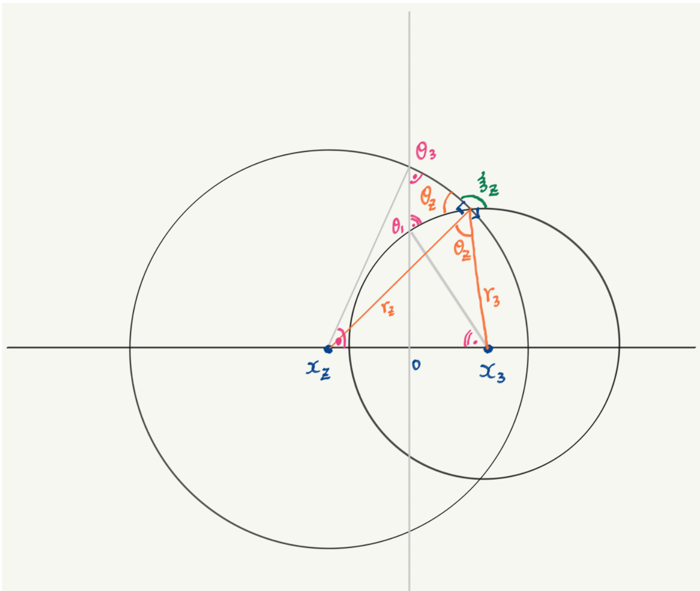

Figure 11.1.3: Computing θ_z

More precisely, we have

$$\begin{aligned}\cos\theta_z &= \frac{-(x_3-x_z)^2+r_z^2+r_3^2}{2r_zr_3}\\ &= \frac{\sin\theta_3\sin\theta_1(y^2+1)-2y\cos\theta_3\cos\theta_1}{2y}\\ &= \frac{\sin\theta_3\sin\theta_1}{2}\left(y+\frac{1}{y}\right)-\cos\theta_3\cos\theta_1\end{aligned}$$

Notice that y is positive and greater or equals to 1, hence $\cos\theta_z$ is strictly monotonically increasing, hence θ_z is strictly monotonically decreasing. Notice that the limit case is when γ_z and γ_3 share one end point in which case, the angle θ_z is 0. □

Remark 11.1.9.
We use the fact that $\mathrm{Isom}(\mathbb{H})$ acts transitively on pairs of distinct points in $\mathbb{H}$ with a same distance.

Corollary 11.1.10
Two triangles are isometric if and only if they have the same interior angles.

Proposition 11.1.11
Two triangles with ideal vertices are isometric to each other if and only if they have the same interior angles.

Proof. Given two triangles with ideal vertices with the same interior angles, we try to find an isometry sending one to the other. The proof is essentially the same as above. We omit it here. □

Remark 11.1.12.
We use the fact that $\mathrm{Isom}(\mathbb{H})$ acts transitively on $\partial\mathbb{H}\times\mathbb{H}$ and on $(\partial\mathbb{H})^{(2)}$.

Corollary 11.1.13
The sum of interior angles of a hyperbolic triangle is strictly smaller than π. Reciprocally, for any three non-negative numbers with sum smaller than π, we have a unique hyperbolic triangle up to isometry.

11.1.3 Ideal triangles

Ideal triangles are used a lot when studying hyperbolic surfaces. We would like to show some facts about it.

Proposition 11.1.14
All ideal triangles are isometric to each other.

Proof. Since triangles are determined by their interior angles. All triangles with interior angles 0, 0 and 0 are isometric to each other. □

On the other hand one may understand this by looking at the vertices. We have shown that $\mathrm{Isom}(\mathbb{H})$ acts transitively on $(\partial\mathbb{H})^{(3)}$ which can also be considered as ideal triangles marked by ordered vertices.

Proposition 11.1.15
Any triangle is contained in one ideal triangle.

Proof. For any triangle Δ, we may use the construction in the proof of Proposition 11.1.8 to get an ideal triangle such that Δ is contained in the interior of the ideal triangle. □

Let Δ be an ideal triangle.

Proposition 11.1.16
For any point $z \in I_1$, its distance to $I_2 \cup I_3$ is bounded by $(\sqrt{2}+1)/3$.

Proof. Without loss of generality, we may consider the ideal triangle with vertices $z_1 = \infty$, $z_2 = 0$ and $z_3 = 1$. By the symmetry of Δ, we only need to consider the point $z = (1/2) + i(1/2)$ and its distance to the side I_3 which is V_0.

Since the distance is realized by the circular geodesic segment with center at 0, we consider the angle bounded by the two Euclidean radius which is $\pi/4$. Hence the distance from z to V_0 is $(\sqrt{2}+1)/3$. □

Corollary 11.1.17
For any triangle, a point on one side is in the δ-neighborhood of the union of the other two sides, for any $\delta > (\sqrt{2}+1)/3$.

Remark 11.1.18.
We only consider open neighborhoods.

Remark 11.1.19.
This is not true for Euclidean triangles, since we have the rescaling to make the distance to be bigger and bigger with no limit. This is one way to say triangles in hyperbolic space are thin. This notion is generalized to the one called *Gromov δ-hyperbolicity*.

11.1.4 Trigonometry formulas

To study the trigonometry geometry of triangles, we will start by looking at some special types, then use them to get information for general triangles. As we discussed before, by using isometry of $\mathbb{H}$, we will renormalize triangles to some standard position when we make the discussion in the following parts.

Triangles with interior angles $(0, \alpha, \alpha)$

We consider a isosceles triangle Δ with one ideal vertex. Let $\alpha \in (0, \pi/2)$. Without loss of generality, we may assume that the vertices of Δ are $z_1 = \infty$, $z_2 = e^{i(\pi-\alpha)}$ and $z_3 = e^{i\alpha}$. Hence z_2 and z_3 are on a same circular geodesic with center 0 and radius $r = 1$.

The finite ones among all side lengths and angles in this case are $\theta_2 = \theta_3 = \alpha$, $l_1 = l$. Using the distance formula, we have the following relation:

$$\begin{aligned} l &= \log \frac{\sin(\pi-\alpha)}{\cos(\pi-\alpha)+1} - \log \frac{\sin\alpha}{\cos\alpha+1} \\ &= \log \frac{1+\cos\alpha}{1-\cos\alpha}. \end{aligned}$$

This can be in turn rewritten as

$$\begin{aligned} \coth\frac{l}{2}\cos\alpha &= 1, \\ \cosh\frac{l}{2}\sin\alpha &= 1, \\ \sinh\frac{l}{2}\tan\alpha &= 1. \end{aligned}$$

We compute the area of Δ. First we consider the area between the vertical geodesics $V_{\cos\alpha}$ and $V_{-\cos\alpha}$ above $H_{\sin\alpha}$. We denote this region by K

$$\begin{aligned} A_{\mathbb{H}}(K) &= \int_{\sin\alpha}^{\infty}\int_{-\cos\alpha}^{\cos\alpha}\frac{1}{y^2}\,\mathrm{d}x\,\mathrm{d}y \\ &= \int_{\sin\alpha}^{\infty}\frac{2\cos\alpha}{y^2}\,\mathrm{d}y \\ &= -\frac{2\cos\alpha}{y}\Big|_{\sin\alpha}^{\infty} \\ &= \frac{2\cos\alpha}{\sin\alpha} = 2\cot\alpha. \end{aligned}$$

Now we consider the region K' which is the intersection between the unit disk and K.

$$\begin{aligned} A_{\mathbb{H}}(K') &= \int_{\sin\alpha}^{1}\int_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}}\frac{1}{y^2}\,\mathrm{d}x\,\mathrm{d}y \\ &= \int_{\sin\alpha}^{1}\frac{2\sqrt{1-y^2}}{y^2}\,\mathrm{d}y \\ &= \int_{\alpha}^{\frac{\pi}{2}}\frac{2\cos t}{\sin^2 t}\,\mathrm{d}\sin t \\ &= 2\int_{\alpha}^{\frac{\pi}{2}}\left(\frac{1}{\sin^2 t}-1\right)\mathrm{d}t \\ &= 2\cot\alpha - (\pi - 2\alpha). \end{aligned}$$

Hence we have

$$A_{\mathbb{H}}(\Delta) = A_{\mathbb{H}}(K) - A_{\mathbb{H}}(K') = \pi - 2\alpha.$$

Triangles with interior angles $(0, \pi/2, \alpha)$

This triangle can be considered as half of a triangle with interior angles 0, α and α. Without loss of generality, we assume that Δ is given by $z_1 = \infty$, $z_2 = i$ and $z_3 = e^{i\alpha}$. Let $l_1 = l$. Using the computation from the previous case, we have

$$\begin{aligned}\coth l \cos\alpha &= 1,\\ \cosh l \sin\alpha &= 1,\\ \sinh l \tan\alpha &= 1.\end{aligned}$$

and

$$A_{\mathbb{H}}(\Delta) = \frac{\pi}{2} - \alpha.$$

Triangles with interior angles $(0, \alpha, \beta)$

Without loss of generality, we may assume that the triangle Δ is given by $z_1 = \infty$, $z_2 = e^{i(\pi-\beta)}$ and $z_3 = e^{i\alpha}$. Let $l_1 = l$.

We consider two cases. The first case is that both α and β are in $(0, \pi/2)$. In this case, we can combine the two triangles to get Δ. The first triangle is given by $w_1 = \infty$, $w_2 = i$ and $w_3 = e^{i\alpha}$, and the second triangle is given by $w_1' = \infty$, $w_2' = i$ and $w_3' = e^{i(\pi-\beta)}$. We denote by t_1 and t_2 the finite sides in the first and second triangles respectively.

Hence we have

$$\begin{aligned}\cosh l &= \cosh t_1 \cosh t_2 + \sinh t_1 \sinh t_2\\ &= \frac{1}{\sin\alpha\sin\beta} + \frac{1}{\tan\alpha\tan\beta}\\ &= \frac{1+\cos\alpha\cos\beta}{\sin\alpha\sin\beta}.\end{aligned}$$

and

$$\begin{aligned}\sinh l &= \sinh t_1 \cosh t_2 + \sinh t_2 \cosh t_1\\ &= \frac{1}{\tan\alpha\sin\beta} + \frac{1}{\tan\beta\sin\alpha}\\ &= \frac{\cos\alpha+\cos\beta}{\sin\alpha\sin\beta}.\end{aligned}$$

And the area of Δ is given by

$$A_{\mathbb{H}}(\Delta) = \left(\frac{\pi}{2} - \alpha\right) + \left(\frac{\pi}{2} - \beta\right) = \pi - \alpha - \beta.$$

The second case is that one of α and β is in $(\pi/2, \pi)$. We may verify that the above formula still hold for this case.

Triangles with interior angles $(\theta_1, \theta_2, \theta_3)$

We may show the following formulas for general triangles:

- The Sine rule:
$$\frac{\sinh l_1}{\sin\theta_1} = \frac{\sinh l_2}{\sin\theta_2} = \frac{\sinh l_3}{\sin\theta_3};$$

- The Cosine rule I:

$$\cosh l_3 = \cosh l_1 \cosh l_2 - \sinh l_1 \sinh l_2 \cos\theta_3,$$

- The Cosine rule II:

$$\cos\theta_3 = \sin\theta_1 \sin\theta_2 \cosh l_3 - \cos\theta_1 \cos\theta_2.$$

The area for a general triangle is given by

$$A_{\mathbb{H}}(\Delta) = \pi - (\theta_1 + \theta_2 + \theta_3).$$

11.2 Polygons

Let P be a general polygon with n sides. Let $\theta_1, ..., \theta_n$ to be the interior angles of P. Using triangulation of P, we can show that

$$A_{\mathbb{H}}(P) = (n-2)\pi - (\theta_1 + \cdots + \theta_n).$$

We may try to study the relations among side lengths and interior angles as what we have done for triangles. We will not list all of them here but only one that we may use a lot later for the right angle hexagon:

- The Sine rule:

$$\frac{\sinh l_1}{\sinh l_4} = \frac{\sinh l_2}{\sinh l_5} = \frac{\sinh l_3}{\sinh l_6},$$

- The cosine rule:

$$\cosh l_3 = \cosh l_1 \cosh l_2 - \sinh l_1 \sinh l_2 \cosh l_4.$$

Remark 11.2.1.
We may verify that

$$i\sin x = \sinh(ix),$$
$$\cos x = \cosh(ix).$$

In some sense, we may consider the length in the right angle hexagon as imaginary angles.

Remark 11.2.2.
We recall some trigonometry formulas from Euclidean geometry:

- The Sine rule:

$$\frac{l_1}{\sin\theta_1} = \frac{l_2}{\sin\theta_2} = \frac{l_3}{\sin\theta_3}$$

- The Cosine rule:

$$l_3^2 = l_1^2 + l_2^2 - 2l_1 l_2 \cos\theta_3.$$

12 Convex sets

12.1 Definition

Definition 12.1.1
A subset of $\mathbb{H}$ is said to be *convex* if for any pair of points in it, the geodesic segment connecting these two points is also contained in this subset.

We say it is *strictly convex* if there is no geodesic segment contained in its boundary.

12.2 Describe convex sets with half planes

To study convex subset, we use a lot half planes in $\mathbb{H}$. Let γ be a complete geodesic. It separate $\mathbb{H}$ into two disconnected parts.

Definition 12.2.1
Each part together with γ is called a half plane associated to γ.

Proposition 12.2.2
Any half plane is convex.

Proof. Since all geodesics belong to either straight line or circles symmetric to the real axis, given any two geodesics they only intersect once. If a half plane is not convex, there will be another geodesic intersecting the boundary with two points, which is a contradiction. □

Corollary 12.2.3
The intersection of any two half planes is convex.

Proof. An intersection among convex sets is convex. □

Let K be a closed convex subset in $\mathbb{H}$. Let $\mathcal{H}_K$ be the collection of all half planes containing K. Then we have the following proposition:

Proposition 12.2.4

$$\cap_{H \in \mathcal{H}_K} H = K.$$

Proof. To show this, we prove the following two things:

(i) $\cap_{H \in \mathcal{H}_K} H \supset K$,

(ii) $\cap_{H \in \mathcal{H}_K} H \subset K$.

By the definition of $\mathcal{H}_K$, we have $K \subset H$ for all $H \in \mathcal{H}_K$, hence we have the first.

For the second, we may consider the complementary of both sets. Let w be a point not in K. We may consider the geodesic ray issued from w parametrized by S^1. Each ray either intersects K or not. We denote by I the subset of S^1 consisting of rays intersecting K.

We claim that I is an interval. Since given any two rays intersecting K, there are two points one on each ray contained in K. By convexity, the geodesic segment connecting these two points will also be in K. Hence one of the segment in S^1 connecting the parameters of the two rays will be contained in I. Since it is path connected, we have I is an interval.

Secondly, the closure of I is strictly contained in the interior of a half circle of S^1. Otherwise, there will be a ray σ and its opposite ray $\bar{\sigma}$ contained in $\bar{I}$. We denote by $e^{i\theta}$ and $-e^{i\theta}$ the parameter for σ and $-\sigma$ respectively. We consider a sequence $(z_n)_n$ of parameters in I converges to $e^{i\theta}$ and a sequence $(w_n)_n$ of parameters in I converges to $-e^{i\theta}$. Hence we get a sequence of $(\gamma_n)_n$ where γ_n is a geodesic intersecting the rays with parameters z_n and w_n where the segment between these two rays are contained in K.

Hence the end points of γ_n converge to the end point of the geodesic given by $\sigma \cup (-\sigma)$. Moreover the points of K on the ray with parameter z_n will converges to a point on $\bar{\sigma}$ and the points of K on rays with parameters w_n will converges to a point on $-\sigma$. For any of these two points, if it is in $\mathbb{H}$, then it must be in K , since K is closed. Otherwise, it is the ending point of the geodesic ray σ or $-\sigma$. In either case, the geodesic segment connecting these two limit points should be at least on the boundary of K. Since K is closed, this geodesic segment will be in K. Since it contains w, then w must be in K as well. Hence a contradiction to the assumption that w is not in K.

We denote by J the half circle which strictly contains $\bar{I}$. Since the complement of K is open, we may choose a ball B_w in centered at w and disjoint from K. The set $J \setminus \bar{I}$ consists of two connected components with non empty interior. We consider one parameter in the interior of each connected component. Consider the two rays associated to these two parameters. We choose two points in the intersection between the two rays and B_w, one for each ray. The geodesic passing through them will not intersect K. At the same time, it separates w from K. Hence the complement of K is also contained in the complement of the intersection among the half spaces $H \in \mathcal{H}_K$. This prove the second point. □

Example 12.2.5.

We give some examples of convex sets: Half plane, circle, triangle, convex polygon.

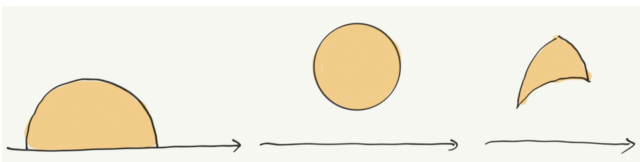

Figure 12.2.1: Convex sets

as well as non examples: non-convex polygon.

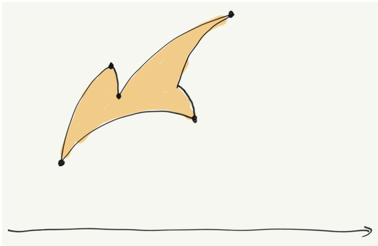

Figure 12.2.2: Non-convex polygon

12.3 Convex curves

Let γ be a path in $\mathbb{H}$, with no self-intersection and separating $\mathbb{H}$ into two connected components H and H'.

Definition 12.3.1
We say that γ is *convex* to H (resp. to H') if H (resp. H') is a convex set in $\mathbb{H}$.

Remark 12.3.2.
In the other word, the curve γ is on the boundary of a convex set.

Proposition 12.3.3
A curve γ is convex to both H and H' if and only if it is a complete geodesic.

Proof. We may repeat the proof above using rays on γ. The maximal interval in S^1 for H and H' have disjoint interiors, since they are disjoint. If one of them is contained in the interior of a half circle, we may find two rays with angle smaller than π bigger than 0, such that the part of $\mathbb{H}$ between them is in the complement of $H \cup H'$ which is contradiction, since $\gamma \cup H \cup H' = \mathbb{H}$. □

Proposition 12.3.4
If one of H and H' is compact, it will be the convex set.

Proof. Since one of them is compact which can be covered by finitely many finite radius balls, then the other one have the entire boundary at infinity as its boundary in $\overline{\mathbb{H}}$. Hence all geodesics should be in that part, which is a contradiction. □

Example 12.3.5.
Below are some examples of convex curves to only one side: circle, horocycle, hypercycle. The orange domain is convex in each picture. The green line and blue line are geodesics to show that they are convex to only one side.

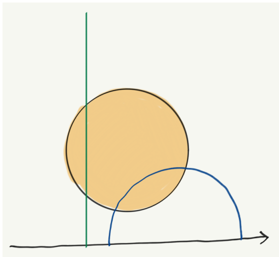

Figure 12.3.1: Circle

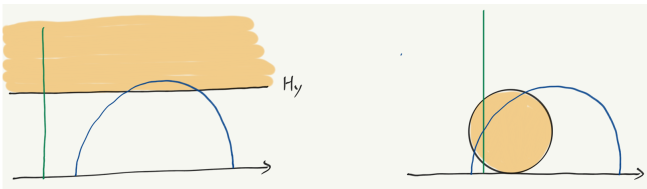

Figure 12.3.2: Horocycles

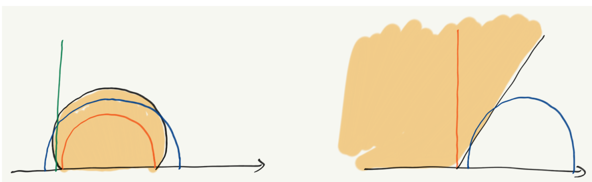

Figure 12.3.3: Hypercycles

Bibliography

[1] Alan F. Beardon, *The geometry of discrete groups*, Graduate Texts in Mathematics, vol. 91, Springer-Verlag, New York, 1983.

[2] Benson Farb and Dan Margalit, *A primer on mapping class groups*, Princeton Mathematical Series, vol. 49, Princeton University Press, Princeton, NJ, 2012.

[3] *Travaux de Thurston sur les surfaces*, Société Mathématique de France, Paris, 1991. Séminaire Orsay, Reprint of ıt Travaux de Thurston sur les surfaces, Soc. Math. France, Paris, 1979 [MR0568308 (82m:57003)], Astérisque No. 66-67 (1991) (1991).

[4] Greg McShane, *A remarkable identity for lengths of curves*, ProQuest LLC, Ann Arbor, MI, 1991. Thesis (Ph.D.)–University of Warwick (United Kingdom).

[5] Svetlana Katok, *Fuchsian groups*, Chicago Lectures in Mathematics, University of Chicago Press, Chicago, IL, 1992.

[6] Maryam Mirzakhani, *Simple geodesics and Weil-Petersson volumes of moduli spaces of bordered Riemann surfaces*, Invent. Math. **167** (2007), no. 1, 179–222.

[7] ———, *Growth of the number of simple closed geodesics on hyperbolic surfaces*, Ann. of Math. (2) **168** (2008), no. 1, 97–125.

[8] R. C. Penner and J. L. Harer, *Combinatorics of train tracks*, Annals of Mathematics Studies, vol. 125, Princeton University Press, Princeton, NJ, 1992.