Notes on Non-Euclidean Geometry

Jiayin Pan

Abstract. Lecture notes for Math 113 Non-Euclidean geometry. We loosely follow the textbook *Geometries and Groups* by Nikulin and Shafarevich. Most proofs have been rewritten and more content has been added. This note is self-contained.

Last updated: December 14, 2020

Contents

Chapter 1. Motivating examples 5
1.1. Sphere 5
1.2. Cylinder 7
1.3. More examples 10

Chapter 2. Basic Concepts 13
2.1. Geometries 13
2.2. Groups 16
2.3. Group actions on Geometries 18

Chapter 3. Free and discrete isometric actions on $\mathbb{R}^2$ or $\mathbb{S}^2$ 23
3.1. Hello again, linear algebra 23
3.2. Elements of $\mathrm{Isom}(\mathbb{R}^2)$ without fixed points 25
3.3. Subgroups of $\mathrm{Isom}(\mathbb{R}^2)$ 26
3.4. Subgroups of $\mathrm{Isom}(\mathbb{S}^2)$ 30

Chapter 4. Coverings 33
4.1. Covering maps 33
4.2. Cover a 2-dimensional locally Euclidean space by $\mathbb{R}^2$ 35
4.3. Covering groups 37

Chapter 5. Hyperbolic Plane 41
5.1. Introduction 41
5.2. Pure imaginary half line under the $SL_2(\mathbb{R})$-action 42
5.3. Distance function on $\mathbb{H}$ 44
5.4. Projection 48
5.5. The isometry group of $(\mathbb{H}, d)$ 51

CHAPTER 1

Motivating examples

1.1. Sphere

We consider the 2-dimensional unit sphere $\mathbb{S}^2$ in $\mathbb{R}^3$ centered at the origin. We would like to measure the distance between two points, say P and Q, on the sphere. One way is using the distance from $\mathbb{R}^3$, but a more natural way (from the perspective of $\mathbb{S}^2$) is the length of a shortest path on $\mathbb{S}^2$ between P and Q. We denote $d_S(P, Q)$ as their distance coming from the length of shortest path.

Definition 1.1.1. A *great circle* on $\mathbb{S}^2$ is the intersection of $\mathbb{S}^2$ and a plane through the origin.

Note that great circle is a circle of radius 1. For $P, Q \in \mathbb{S}^2$ and a great circle through them, P and Q divide the circle into two arcs. When speaking of the arc between P and Q, we mean the shorter one unless otherwise noted. Note that this arc is unique unless $P = -Q$.

Theorem 1.1.2. *The shortest path on $\mathbb{S}^2$ between $P, Q \in \mathbb{S}^2$ is the arc of a great circle through P and Q.*

The proof is a contradicting argument. Before presenting the detailed proof, here is a sketch:
1. Suppose the contrary, then there is a path $\tilde{l}$ between P, Q that is shorter than the arc l.
2. Choose a series of points on $\tilde{l}$, then connecting them successively by small arcs. In this way, we obtain a "broken arc" l' between P, Q. As a technical part, we will show that with suitable intermediate points chosen, l' has length less than that of l.
3. Deleting the intermediate points one by one, while connecting the remaining points by arcs. In this process, the new broken arc has length no greater than that of the old broken arc (Surely, we will need a lemma here).
4. Eventually, after all the intermediate points are deleted, we end in a arc between P and Q, but its length is shorter than the arc that we start with. A contradiction.

Now we prove some lemmas to address the non-trivial part in the sketch.

Lemma 1.1.3. *Let $A, B \in \mathbb{S}^2$. Let d be the Euclidean distance between them and $\theta = \angle AOB$. Then*

$$d = \theta - O(\theta^3),$$

where $O(\theta^3)$ means a term comparable to θ^3 as $\theta \to 0$.

Proof. By cosine law, we have

$$1^2 + 1^2 - 2 \cdot 1 \cdot \cos\theta = d^2,$$

or $d^2 = 2 - 2\cos\theta$. When θ is very small, we can apply the Maclaurin series

$$d^2 = 2 - 2\cos\theta = 2 - 2\left(1 - \frac{\theta^2}{2} + O(\theta^4)\right) = \theta^2 - O(\theta^4).$$

Thus

$$d = \theta\sqrt{1 - O(\theta^2)} = \theta(1 - O(\theta^2)) = \theta - O(\theta^3).$$

(You may check the Maclaurin series of $\sqrt{1+x}$ if you are not sure about the second equality above.) □

Remarks:
(1) $\theta = \angle AOB$ is equal to the length of the arc between A and B.
(2) $d \leq d_S(A, B) \leq \theta$.

Lemma 1.1.4. *Let $P, Q \in \mathbb{S}^2$ and let $\tilde{l}$ be a path between them. Let $P_0 = P, P_1, ..., P_n = Q$ on $\tilde{l}$ such that they divide the curve $\tilde{l}$ evenly. Let l_n be the broken arc through $P_0, ..., P_n$. Then*

$$\text{Length}(l_n) - \text{Length}(\tilde{l}) \to 0$$

as $n \to \infty$.

Proof. By assumptions, we have

$$d_S(P_i, P_{i+1}) = \text{Length}(\tilde{l})/n =: \delta(n)$$

for each $i = 0, 1, ..., n-1$. Since the sphere is soooo symmetric, the angle between $\overrightarrow{OP_i}$ and $\overrightarrow{OP_{i+1}}$ are the same for all i; we denote this angle as $\theta(n)$. Then

$$\text{Length}(l_n) = n\theta(n), \quad \text{Length}(\tilde{l}) = n\delta(n) \geq nd(n),$$

where $d(n)$ is the Euclidean distance between P_i and P_{i+1}. By Lemma 1.1.3, we estimate

$$\begin{aligned} \text{Length}(l_n) - \text{Length}(\tilde{l}) &= n(\theta(n) - \delta(n)) \leq n(\theta(n) - d(n)) \\ &= n \cdot O(\theta(n)^3) = \frac{\text{Length}(\tilde{l})}{\delta(n)} \cdot O(\theta(n)^3) \\ &\leq \text{Length}(\tilde{l}) \cdot \frac{\theta(n)}{d(n)} \cdot O(\theta(n)^2). \end{aligned}$$

Since $\theta(n)/d(n) \to 1$ and $\theta(n) \to 0$ as $n \to \infty$, we see that the error above $\to 0$ as $n \to \infty$. □

Lemma 1.1.5. *Let $A, B, C \in \mathbb{S}^2$. Then*

$$\angle AOC \leq \angle AOB + \angle BOC.$$

Proof. We project B to $p(B)$ on plane AOC. It suffices to show that

$$\angle AOp(B) \leq \angle AOB, \quad \angle p(B)OC \leq \angle BOC.$$

We prove the first inequality (the second one is the same). Without lose of generality, we assume that $A = (1, 0, 0)$ and plane AOC is the xy-plane. We write $\vec{v} = \vec{OA} = (1, 0, 0)$, $\vec{w} = \vec{OB} = (x, y, z)$. Then $p(\vec{w}) = O\vec{p(B)} = (x, y, 0)$. We have

$$x = \vec{v} \cdot \vec{w} = \cos\angle AOB, \quad x = \vec{v} \cdot p(\vec{w}) = ||p(w)|| \cos\angle AOp(B).$$

Since $||p(w)|| \leq 1$, we conclude that $\angle AOp(B) \leq \angle AOB$. □

PROOF OF THEOREM 1.1.2. We follow the sketch. Let l be an arc between P, Q. Suppose that there is a path $\tilde{l}$ on $\mathbb{S}^2$ between P, Q such that

$$\text{length}(\tilde{l}) < \text{length}(l).$$

Let $P_0 = P, P_1, ..., P_n = Q$ be a series of points of $\tilde{l}$ that divide $\tilde{l}$ evenly. With these points, we construct a broken arc, by connecting P_i to P_{i+1} successively by an arc. We call this broken arc l_n. By Lemma 1.1.4, we can assume that n is sufficiently large such that

$$\text{length}(l_n) < \text{length}(l).$$

Next we delete the intermediate points $P_1, ..., P_{n-1}$ one by one. First delete P_1, that is, we replace two arcs, P_0 to P_1 and P_1 to P_2, by a new arc from P_0 directly to P_2. By Lemma 1.1.5, the length of the new arc is no greater than the sum of two original orcs (recall that the angle equals the length of arc). We call the new broken arc through $P_0, P_2, ..., P_n$ as $l_{n,1}$. Then

$$\text{length}(l_{n,1}) \leq \text{length}(l_n) < \text{length}(l).$$

We continue this process until all intermediate points are deleted. We end in an arc $l_{n,n-1}$ between P, Q with

$$\text{length}(l_{n,n-1}) < \text{length}(l).$$

This is a contradiction to our choice l as the shortest arc between P and Q. □

1.2. Cylinder

We look for shortest paths on the cylinder $C = S^1 \times \mathbb{R}$, as a surface in $\mathbb{R}^3$.

We write a point on C as (θ, z), which corresponds to the point $(\cos\theta, \sin\theta, z)$ in $\mathbb{R}^3$. Given two distinct points $P = (\theta_1, z_1)$ and $Q = (\theta_2, z_2)$ on C, we define an *arc* connecting them as below:
(1) if $\theta_1 = \theta_2$, then the arc is the vertical segment connecting P and Q;
(2) if $z_2 = z_2$, then the arc is the usual arc (as part of a circle) lying in the horizontal plane containing P and Q;
(3) if $\theta_1 \neq \theta_1$ and $z_2 \neq z_2$, we define the arc as part of a helix:

$$(\alpha(t), z_1 + t(z_2 - z_1)), \quad t \in [0, 1],$$

where $\alpha(t)$ is the usual arc from θ_1 to θ_2 in the circle.

Our goal is showing that these arcs are exactly all the shortest curves on C. We first calculate the length of the arcs. Let P and Q be two points on the cylinder. Using the symmetries of the cylinder, we can assume that $P = (0, 0)$ and $Q = (\theta, z)$ with $\theta \in [0, \pi]$. For cases (1) and (2) above, the length of the corresponding arc is clear. For case (3), the arc has parametric equation

$$\gamma(t) = (\cos(\theta t), \sin(\theta t), tz), \quad t \in [0, 1].$$

It has tangent vector

$$\gamma'(t) = (-\theta \sin(\theta t), \theta \cos(\theta t), z)$$

with norm

$$||\gamma'(t)|| = \sqrt{\theta^2 + z^2}.$$

Thus the arc has length

$$\text{Length}(\gamma) = \int_0^1 ||\gamma'(t)|| dt = \sqrt{\theta^2 + z^2}.$$

Once we know that these arcs are indeed the shortest curves, the distance between $P = (0, 0)$ and $Q = (\theta, z)$ will be $\sqrt{\theta^2 + z^2}$, where $\theta \in [0, \pi]$. Note that this formula is identical to the Pythagorean Theorem of the plane. For $P = (0, 0)$ and $Q = (\theta, z)$ with $\theta \in (\pi, 2\pi)$, we have

$$d(P, Q) = d(P, Q') = \sqrt{(2\pi - \theta)^2 + z^2},$$

where $Q' = (2\pi - \theta, z)$

Theorem 1.2.1. *The shortest path between P, Q on C is the arc between them.*

To show that these arcs are indeed the shortest curves, we follow a similar strategy used in the sphere case.

Exercise 1.2.2. *Write a proof based on the sketch below:*
Step 1. Suppose that there is a secret path whose length is shorter than the arc, then we construct a path by joining small broken arcs; show that the length this new path can be arbitrary close (, as the partition becomes finer,) to that of the secret path.
Step 2. Similar to the sphere case, for the broken arcs, show that after deleting a midpoint, the new broken arc is shorter than the old one.
Step 3. Derive a contradiction.

You may have a try on your own before taking a look at the hints below.

Hint on Step 1. You may encounter with an issue at some point. Though the cylinder is so symmetric, it is not *soooo* symmetric as the sphere. More precisely, once we know the spherical distance between two points, their Euclidean distance is determined (independent of the actual positions of these two points). This property is no longer true on the cylinder. We write $\mathcal{A}(A, B)$ as the length of the arc between A and B; we also write $d_E(A, B)$ as the Euclidean distance between A, B. Inspecting the proof in the spherical case, here we would similarly estimate

$$\begin{aligned}\text{Length}(l_n) - \text{Length}(\tilde{l}) &= \sum_i \mathcal{A}(P_i, P_{i+1}) - n\delta(n) \\ &= \sum_i (\mathcal{A}(P_i, P_{i+1}) - \delta(n)) \\ &\leq \sum_i (\mathcal{A}(P_i, P_{i+1}) - d_E(P_i, P_{i+1}))\end{aligned}$$

In the sphere case, each arc has the same length $\theta(n)$ and the Euclidean distance are also the same $d(n)$. On the cylinder, these two terms may depend on the position of points. Fortunately, it suffices to have a estimate: can you find a useful upper bound for $\mathcal{A}(P_i, P_{i+1}) - d_E(P_i.P_{i+1})$?

Hint on Step 2. This is indeed the triangle inequality of the plane (Why?)

Besides treating cylinder as a surface in $\mathbb{R}^3$, we can define it by using a equivalence relation.

Definition 1.2.3. For two points $A = (x_1, y_1)$ and $B = (x_2, y_2)$, we write $x \sim y$, if $x_1 - x_2 = 2k\pi$ for some $k \in \mathbb{Z}$ and $y_1 = y_2$.

It is straight-forward to check that $\sim$ is an equivalence relation (Have a try if you are not sure).

Equivalently, we can introduce a map $\phi : \mathbb{R}^2 \to \mathbb{R}^2$, $\phi(x, y) = (x+2\pi, y)$. Then $x \sim y$ if and only if $y = \phi^k(x)$ for some $k \in \mathbb{Z}$.

Let $E(A)$ be the set of all points that is equivalent to A, that is,

$$E(A) = \{A' \in \mathbb{R}^2 | A' = \phi^k(A) \text{ for some } k \in \mathbb{Z}\}.$$

We write $\bar{A}$ and $\bar{B}$ as the corresponding points in the quotient set $\mathbb{R}^2/\sim$. We define a distance function $\bar{d}$ on $\mathbb{R}^2/\sim$ as follows:

$$\bar{d}(\bar{A}, \bar{B}) = d(E(A), E(B)) = \inf_{A' \in E(A), B' \in E(B)} d(A', B').$$

Remark: If you never see inf before, just treat it as minimum.

Proposition 1.2.4. *The above infimum(minimum) can be achieved. Also, there is* $B_0 \in E(B)$ *such that*

$$\bar{d}(\bar{A}, \bar{B}) = d(A, B_0).$$

PROOF. Note that for any $A' = \phi^k(A) \in E(A)$,

$$\begin{aligned} d(A', E(B)) &= d(\phi^k(A), E(B)) = d(\phi^{-k}\phi^k(A), \phi^{-k}(E(B))) \\ &= d(A, \phi^{-k}(E(B))) = d(A, E(B)). \end{aligned}$$

The second equality holds because ϕ^{-k}, as a translation, preserves the distance function. Thus

$$d(E(A), E(B)) = d(A, E(B)) = \min_{B' \in E(B)} d(A, B').$$

Choose $B_0 \in E(B)$ such that $d(A, E(B)) = d(A, B')$. The result follows. □

It is straight-forward to check that $\bar{d}$ satisfies:
(1) $\bar{d}(\bar{A}, \bar{B}) \geq 0$, $=$ holds if and only if $\bar{A} = \bar{B}$;
(2) $\bar{d}(\bar{A}, \bar{B}) = \bar{d}(\bar{B}, \bar{A})$;
(3) $\bar{d}(\bar{A}, \bar{B}) \leq \bar{d}(\bar{A}, \bar{C}) + \bar{d}(\bar{C}, \bar{B})$.

We give details on (3). Applying Proposition 1.2.4, we can find $A_0, B_0, C_0 \in \mathbb{R}^2$ such that

$$\bar{d}(\bar{A}, \bar{C}) = d(A_0, C_0), \quad \bar{d}(\bar{C}, \bar{B}) = d(C_0, B_0).$$

Then

$$\bar{d}(\bar{A}, \bar{B}) \leq d(A_0, B_0) \leq d(A_0, C_0) + d(C_0, B_0) = \bar{d}(\bar{A}, \bar{C}) + \bar{d}(\bar{C}, \bar{B}).$$

The next proposition explains how the "strip with portals" model is related to the "quotient space" model.

Proposition 1.2.5. *There is a one-to-one correspondence between* $\mathbb{R}^2/\sim$ *(the quotient space model) and the half-open-half-closed stripe* $[0, 2\pi) \times \mathbb{R}$.

PROOF. We define

$$F : [0, 2\pi) \times \mathbb{R} \to \mathbb{R}^2/\sim (x, y) \mapsto \overline{(x, y)},$$

where $\overline{(x, y)}$ is the corresponding point in the quotient set.

F is surjective (onto): For any $\overline{(x, y)} \in \mathbb{R}^2/\sim$, since $x \in \mathbb{R}$, we write x as $x = 2k\pi + x'$ for some $k \in \mathbb{Z}$ and $x' \in [0, 2\pi)$. Note that $(x, y) \sim (x', y)$. Thus

$$F((x', y)) = \overline{(x', y)} = \overline{(x, y)}.$$

F is injective (one-to-one): Suppose that $F(x_1, y_1) = F(x_2, y_2)$, where (x_1, y_1), $(x_2, y_2) \in [0, 2\pi) \times \mathbb{R}$. By the definition of F, this means $\overline{(x_1, y_1)} = \overline{(x_2, y_2)}$. Thus $(x_1, y_1) \sim (x_2, y_2)$, that is,

$$x_1 - x_2 = 2k\pi, \quad y_1 = y_2,$$

for some $k \in \mathbb{Z}$. Recall that $x_1, x_2 \in [0, 2\pi)$. It follows that k must be zero. Hence $(x_1, y_1) = (x_2, y_2)$. □

Let $\bar{A} \in \mathbb{R}^2/\sim$. We write $B_r(\bar{A})$ as the (open) r-ball around $\bar{A}$, that is,

$$B_r(\bar{A}) = \{\bar{P} \in \mathbb{R}^2/\sim |\bar{d}(\bar{A}, \bar{P}) < r\}.$$

Proposition 1.2.6. *Let $A \in \mathbb{R}^2$ and $\bar{A} \in \mathbb{R}^2/\sim$. Then $B_{\pi/2}(\bar{A})$ is "identical" to $B_{\pi/2}(A)$ in $\mathbb{R}^2$, in the sense that, for any $B, C \in B_{\pi/2}(A)$, we have $\bar{d}(\bar{B}, \bar{C}) = d(B, C)$.*

Here "identical" means the distance functions are the same. The identification is through the quotient map

$$p : \mathbb{R}^2 \to \mathbb{R}^2/\sim, \quad A \mapsto \bar{A}.$$

Later in the course, we will introduce a rigorous definition to describe this.

Proof. We first show that if $d(A, B) < \pi$, then $d(A, B) = \bar{d}(\bar{A}, \bar{B})$. To prove this, we draw two vertical lines in $\mathbb{R}^2$: The first one l_1 has equal distance between A and $\phi(A)$; the second one l_2 has equal distance between A and $\phi^{-1}(A)$. Since ϕ is a translation along x-axis by 2π, each line has distance π to A. Since $d(A, B) < \pi$, B must stay inside the strip formed by l_1 and l_2. For convenience, we call this strip S. It is clear that

$$\bar{d}(\bar{A}, \bar{B}) = \min\{d(A, B), d(\phi(A), B), d(\phi^{-1}(A), B)\}$$

since all the other points equivalent to A is further away from B. Because B stays in the strip S, the above minimum is realized at A, that is,

$$d(A, B) = \bar{d}(\bar{A}, \bar{B}).$$

Now for any $B, C \in B_{\pi/2}(A)$, note that

$$d(B, C) \leq d(B, A) + d(A, C) < \pi/2 + \pi/2 = \pi.$$

Thus the result follows from what we proved in the first paragraph. □

Remark: When B is in the strip S above, then $d(A, B) = \bar{d}(\bar{A}, \bar{B})$.

1.3. More examples

Torus: We define an equivalence relation on $\mathbb{R}^2$ as follows: we write $(x_1, y_1) \sim (x_2, y_2)$, if there are integers k and l such that

$$x_1 - x_2 = 2k\pi, \quad l_1 - l_2 = 2l\pi.$$

It is direct to check that $\sim$ is indeed an equivalence relation.

The other way to say this is the following. We define two motions on $\mathbb{R}^2$. They are $\phi : (x, y) \mapsto (x + 2\pi, y)$ and $\psi : (x, y) \to (x, y + 2\pi)$. Then $(x_1, y_1) \sim (x_2, y_2)$ if and only if there are integers k and l such that

$$(x_1, y_1) = (\phi^k \circ \psi^l)(x_2.y_2).$$

We define the quotient distance on $\mathbb{R}^2/\sim$ as usual. We denote the quotient distance as d_T. Similar to the cylinder case, we can prove the propositions below.

Proposition 1.3.1. *There is a one-to-one correspondence between* $\mathbb{R}^2/\sim$ *and the half-open-half-closed square* $[0, 2\pi) \times [0, 2\pi)$.

This allows us to write any point in the torus as (α, β), where $\alpha, \beta \in [0, 2\pi)$.

Proposition 1.3.2. *For any* $\bar{A} \in \mathbb{R}^2/\sim$*,* $B_{\pi/2}(\bar{A})$ *is "identical" to* $B_{\pi/2}(A)$ *in* $\mathbb{R}^2$*, in the sense that, for any* $B, C \in B_{\pi/2}(A)$*, we have* $d_T(\bar{B}, \bar{C}) = d(B, C)$.

Exercise 1.3.3. *Prove the above two propositions.*

We have seen a similar property for cylinders and twisted cylinders: local balls in the space are identical to local balls in $\mathbb{R}^2$. We will call this property *locally Euclidean*.

Proposition 1.3.2 also provides a formula for the distance d_T between $(0, 0)$ and (α, β), given that α and β are small:

$$d_T((0, 0), (\alpha, \beta)) = \sqrt{\alpha^2 + \beta^2}.$$

Twisted cylinder (Mobius strip): A glide reflection is a motion of $\mathbb{R}^2$ defined as follows: reflect along a line l, then translate along l. If we set l as the x-axis, we can write a glide reflection as $\psi : (x, y) \to (x + L, -y)$, where L is the length of the translation. For convenience, we will use $L = 2\pi$. Since both reflection and translation preserve the distance function of $\mathbb{R}^2$, so does ψ.

We define an equivalence relation and a quotient distance similarly:

Definition 1.3.4. For $A, B \in \mathbb{R}^2$, we say that $A \sim B$, if there is an integer k such that $B = \psi^k(A)$.

Definition 1.3.5. Distance function on $\mathbb{R}^2/\sim$:

$$\bar{d}(\bar{A}, \bar{B}) = \min_{A' \in E(A), B' \in E(B)} d(A', B').$$

We call the quotient space (with the quotient distance) $(\mathbb{R}^2/\sim, \bar{d})$ a twisted cylinder. Proposition 1.1 and the triangle inequality of $\bar{d}$ extends directly for the twisted cylinder, since we only need the fact ϕ preserves the distance function in the proof of the cylinder case.

Exercise 1.3.6. *Similar to Proposition 1.2.5, show that there is a one-to-one correspondence between the twisted cylinder and the half-open-half-closed stripe* $[0, 2\pi) \times \mathbb{R}$.

Proposition 1.3.7. *For any* $\bar{A} \in \mathbb{R}^2/\sim$*,* $B_{\pi/2}(\bar{A})$ *is "identical" to* $B_{\pi/2}(A)$ *in* $\mathbb{R}^2$*, in the sense that, for any point* $B, C \in B_{\pi}(A)$*, we have* $\bar{d}(\bar{B}, \bar{C}) = d(B, C)$.

Proof. The proof is similar to the cylinder case, with some modifications. It suffices to show that $d(A, B) = \bar{d}(\bar{A}, \bar{B})$ if $d(A, B) < \pi$. For $A \in \mathbb{R}^2$, we draw four lines in $\mathbb{R}^2$. l_1 has equal distance between A and $\phi(A)$, l_2 between A and $\phi^{-1}(A)$, l_3 between A and $\phi^2(A)$, l_4 between A and $\phi^{-2}(A)$. Let R be the region containing A and bounded by these four lines. Note that $d(A, B) < \pi$ implies that $B \in R$. Also, we have

$$d(A, B) = \min_{A' \in E(A)} d(A', B)$$

Thus

$$\bar{d}(\bar{A}, \bar{B}) = \min_{A' \in E(A)} d(A', B) = d(A, B).$$

□

We have seen many examples coming from the quotient of $\mathbb{R}^2$. Here we list two more examples. In each example, we define an equivalence relation on some space via certain motion, then define the quotient distance as usual.

- Cone: let ϕ be the rotation by 90 degrees around the origin. Note that ϕ preserves the distance and has order 4, that is ϕ^4 is the identity map. We define $A \sim B$ if $A = \phi^k(B)$ for some integer k. Note that the quotient space has a cone tip, which is not locally identical to a ball in $\mathbb{R}^2$. Also, this cone tip corresponds to the origin in the plane.
- Closed half-plane: We know this is a subset of $\mathbb{R}^2$, but it can also be realized as a quotient. Let ϕ be reflection about the x-axis. ϕ has order 2. We define $A \sim B$ if $A = \phi^k(B)$ for some integer k (in fact, $k = 0$ or 1). The quotient space has a "boundary part"; any point on the boundary is not locally identical to a ball in $\mathbb{R}^2$.

Besides constructing new geometries from $\mathbb{R}^2$, we can apply a similar idea to the sphere $\mathbb{S}^2$.

Real projective plane $\mathbb{R}P^2$: We define $\phi : S^2 \to S^2$ by sending (x, y, z) to $(-x, -y, -z)$. ϕ has order 2 and does not have fixed points on the sphere. For points on the sphere, we define $A \sim B$ if $A = \phi^k(B)$ for some k (in other words, $A = B$ or $A = \phi(B)$). The quotient space $(S^2/\sim, \bar{d})$ is called the real projective space.

Proposition 1.3.8. *$\mathbb{R}P^2$ is locally spherical.*

SKETCH. For a point A in the sphere, we need to find a region R around A such that the following holds: for any $B \in R$, we have $\bar{d}(\bar{A}, \bar{B}) = d_S(A, B)$. Note that by definition,

$$\bar{d}(\bar{A}, \bar{B}) = \min\{d_S(A, B), d_S(\phi(A), B)\}.$$

Therefore, we only need to make sure that B is closer to A than to $\phi(A)$. Let C be the great circle with equal distance to A and $\phi(A)$. We can set R to be region containing A bounded by this great circle C. Then

$$d_S(A, B) = \bar{d}(\bar{A}, \bar{B})$$

for any $B \in R$. □

CHAPTER 2

Basic Concepts

2.1. Geometries

Definition 2.1.1. Let X be a nonempty set. A *distance function* (or a *metric*) on X is a function $d : X \times X \to [0, \infty)$ satisfying
(1) $d(x, y) = 0$ if and only if $x = y$;
(2) $d(x, y) = d(y, x)$ for all $x, y \in X$;
(3) $d(x, y) \leq d(x, z) + d(z, y)$ for all $x, y, z \in X$.
We call (X, d) a *metric space.*

Examples 2.1.2.
• (Euclidean metric) Standard distance function d_0 on $\mathbb{R}$, or $\mathbb{R}^n$.
• Sphere with the distance defined as the length of the shortest curve.
• Quotient distance that we have seen in previous lectures.
• For a metric space (X, d) and a subset $S \subset X$, we can restrict d on S to obtain a metric on S.
• (Discrete metric) Define d on a set X by setting $d(x, y) = 1$ if $x \neq y$, $d(x, y) = 0$ if $x = y$.
• (Manhattan metric) d on $\mathbb{R}^2$ defined as

$$d((x_1, y_1), (x_2, y_2)) = |x_1 - x_2| + |y_1 - y_2|.$$

• (C^0-metric) We denote $C[0, 1]$ as the set of all continuous functions on $[0, 1]$. Define d on $C[0, 1]$ by

$$d(f, g) = \max_{x \in [0,1]} |f(x) - g(x)|.$$

• Define a non-Euclidean metric d on $\mathbb{R}_+ = (0, \infty)$ by setting

$$d(x, y) = |\ln(x/y)|.$$

As a practice, we show that the last d above is indeed a distance function:
(1) $|\ln(x/y)| = 0$ if and only if $x/y = 1$ if and only if $x = y$.
(2) $\ln(x/y)$ and $\ln(y/x)$ only differs by a sign. Thus $d(x, y) = d(y, x)$.
(3) $d(x, y) = |\ln(x/y)| = |\ln x - \ln y| \leq |\ln x - \ln z| + |\ln z - \ln y|$
$= |\ln(x/z)| + |\ln(z/y)| = d(x, z) + d(z, y).$

Definition 2.1.3. A *curve* (or a *path*) on (X, d) is a continuous map $\gamma : I \to X$, where I is an interval of $\mathbb{R}$.

We do not need the definition of *continuity* anyway, so you can just use your intuition. If you are curious, here is the rigorous definition.

Definition 2.1.4. A map $\gamma : I \to (X, d)$ is *continuous*, if for any $a \in I$ and for any $\epsilon > 0$, there is $\delta > 0$ such that

$$d(\gamma(a), \gamma(t)) \leq \epsilon$$

for all t with $|t-a| \leq \delta$.

Whether a map is continuous depends on d. For example $\gamma(t) = t, t \in [0,1]$ is not continuous in $\mathbb{R}$ with the discrete metric.

Definition 2.1.5. Let $\gamma : [a,b] \to (X,d)$ be a curve on (X,d). Let π be a partition of $[a,b]$, that is, $\pi = \{t_0 = a, t_1, ..., t_k, ...t_n = b\}$. We define the *length* of γ as

$$\lim_{||\pi|| \to 0} \sum_{i=0}^{n-1} d(\gamma(t_i), \gamma(t_{i+1})),$$

(if the limit exists), where $||\pi|| = \max_i |t_{i+1} - t_i|$.

The length of a path could be infinity in general. As a direct consequence of the triangle inequality, the distance between two endpoints is always no greater than the length of γ (As an exercise, you can have a try to prove this).

Definition 2.1.6. (X,d) is called a *length metric space*, if for any pair of points $x, y \in X$, there is a curve γ connecting them with

$$d(x,y) = \text{Length}(\gamma).$$

We call this γ a *minimal geodesic* between x and y.

Some literature may have a different name for Definition 2.1.6, for example, geodesic space, intrinsic metric space.

Example 2.1.7.
- Sphere with the metric defined as the length of shortest curves.
- Any convex subset of $(\mathbb{R}^n, d_0)$ is a length metric space.
- $\mathbb{R}^2 - \{0\}$ is not a length metric space.

As a practice, we show that the last d in Examples 2.1.2 is a length metric space. For $x, y \in \mathbb{R}$, without lose of generality, we may assume that $x < y$. We consider the path

$$\gamma(t) = x + t(y-x), \quad t \in [0,1].$$

Let $\pi = \{t_0 = 0, t_1, ..., t_k, ...t_n = 1\}$ be a partition of $[0,1]$ and let $p_i = \gamma(t_i)$. Note that $p_i < p_{i+1}$. Then

$$\sum_{i=0}^{n-1} d(p_i, p_{i+1}) = \sum_{i=0}^{n-1} |\ln(p_i/p_{i+1})| = \sum_{i=0}^{n-1} (\ln p_{i+1} - \ln p_i) = \ln p_n - \ln p_0 = d(x,y).$$

Since this is true for any partition π, we conclude that $\text{Length}(\gamma) = d(x,y)$.

Definitions 2.1.8. Let (X,d) be a length metric space and let I be an interval. A curve $\gamma : I \to X$ is a *geodesic* if it is locally minimizing, that is, for any $t \in I$, there is $\epsilon > 0$ such that $\text{Length}(\gamma|_{[t-\epsilon, t+\epsilon]}) = d(\gamma(t-\epsilon), \gamma(t+\epsilon))$.

Examples 2.1.9. The followings are some examples of geodesics.
- A great circles in the sphere.
- A helix in the cylinder.

Definition 2.1.10. Let (X, d_X) and (Y, d_Y) be two metric spaces. An *isometry* between X and Y is a bijection $f : X \to Y$ such that

$$d_X(x_1, x_2) = d_Y(f(x_1), f(x_2))$$

for all $x_1, x_2 \in X$. If there is such a map f, we say that X and Y are *isometric.*

If two metric spaces are isometric, we regard them as the same thing. This principle applies to length.

Proposition 2.1.11. *Let $f : X \to Y$ be an isometry between two metric spaces. Let $\gamma : [a,b] \to X$ be a curve in X. Then* $\text{length}(\gamma) = \text{length}(f \circ \gamma)$.

Proof. We denote $\pi = \{t_0 = a, t_1, ..., t_k, ...t_n = b\}$ as a partition of $[a,b]$.

$$\begin{aligned}\text{length}(f \circ \gamma) &= \lim_{||\pi|| \to 0} \sum_{i=0}^{n-1} d_Y(f \circ \gamma(t_i), f \circ \gamma(t_{i+1})) \\ &= \lim_{||\pi|| \to 0} \sum_{i=0}^{n-1} d_X(\gamma(t_i), \gamma(t_{i+1})) \\ &= \text{length}(\gamma).\end{aligned}$$

□

Definition 2.1.12. *(Open) Metric balls.* For $x \in X$ and $r > 0$, we define

$$B_r(x) = \{y \in X \mid d(x,y) < r\}.$$

Definition 2.1.13. Let (X,d) be a length metric space. We say that (X,d) is *2-dimensional locally Euclidean*, if there is $r > 0$ such that for any $x \in X$, $B_r(x)$ is isometric to the r-ball in the Euclidean plane $(\mathbb{R}^2, d_0)$.

Similarly, we can define *2-dimensional locally spherical spaces.*

Remark 2.1.14. $r > 0$ is uniform (works for all $x \in X$) in the above definition. If we do not require this, then any open subset of $\mathbb{R}^2$ would satisfies the property.

Recall that locally Euclidean spaces can be constructed from an equivalence relation in $\mathbb{R}^2$. In these cases, we have seen that the quotient map $p : \mathbb{R}^2 \to \mathbb{R}^2/\sim$ gives an isometry from $B_r(A)$ to $B_r(p(A))$ for $r > 0$ small and any $A \in \mathbb{R}^2$.

Proposition 2.1.15. *Let $(\mathbb{R}^2/\sim, \bar{d})$ is as above. Let γ be a straight line in $\mathbb{R}^2$. Then*
(1) $p \circ \gamma$ is a geodesic in $\mathbb{R}^2/\sim$.
(2) $\text{length}(\gamma|_I) = \text{length}((p \circ \gamma)|_I)$ *for any closed and bounded interval $I \subset \mathbb{R}$.*

Proof. (1) Reparametrize γ is necessary, we assume that it has unit speed, that is

$$d(\gamma(s), \gamma(t)) = |t - s|$$

for all $t, s \in \mathbb{R}$. (For example, you can write $\gamma(t) = v + tw$, where $v \in \mathbb{R}^2$ and w is a unit vector.)

Let $r > 0$ such that $p : B_r(A) \to B_r(p(A))$ is an isometry for all $A \in \mathbb{R}^2$. We pick $\epsilon = r/2$. For any $t \in \mathbb{R}$, we have

$$\begin{aligned}\text{length}(p \circ \gamma|_{[t-\epsilon, t+\epsilon]}) &= \text{length}(\gamma|_{[t-\epsilon,t+\epsilon]}) \\ &= d(\gamma(t-\epsilon), \gamma(t+\epsilon)) \\ &= \bar{d}(p \circ \gamma(t-\epsilon), p \circ \gamma(t+\epsilon)).\end{aligned}$$

The second equality above holds because $\gamma|[t-\epsilon, t+\epsilon]$ has image in $B_r(\gamma(t))$:

$$d(\gamma(s), \gamma(t)) = |t - s| \leq \epsilon < r$$

for all $t \in [t-\epsilon, t+\epsilon]$. This shows that $p \circ \gamma|_{[t-\epsilon,t+\epsilon]}$ is a minimal geodesic for all t. Thus $p \circ \gamma$ is a geodesic.

(2) We again assume that γ is of unit speed. Let $\pi = \{t_0, ..., t_i, ..., t_n\}$ be a partition of I. When $||\pi|| < r$, we have

$$d(\gamma(t_i), \gamma(t_{i+1})) = |t_i - t_{i+1}| < r.$$

Since p is an isometry from $B_r(\gamma(t_i))$ to its image, we see that

$$d(\gamma(t_i), \gamma(t_{i+1})) = d(p \circ \gamma(t_i), p \circ \gamma(t_{i+1})).$$

It follows that

$$\begin{aligned} \text{length}(p \circ \gamma|_I) &= \lim_{||\pi|| \to 0} \sum_{i=0}^{n-1} d(p \circ \gamma(t_i), p \circ \gamma(t_{i+1})) \\ &= \lim_{||\pi|| \to 0} \sum_{i=0}^{n-1} d(\gamma(t_i), \gamma(t_{i+1})) \\ &= \text{length}(\gamma|_I). \end{aligned}$$

□

Corollary 2.1.16. *Let $(\mathbb{R}^2/\sim, \bar{d})$ is as above. Then $(\mathbb{R}^2/\sim, \bar{d})$ is a length metric space.*

Proof. Let $\bar{A}, \bar{B} \in \mathbb{R}^2/\sim$. We know that we can find $A' \in E(A)$ and $B' \in E(B)$ such that

$$\bar{d}(\bar{A}, \bar{B}) = d(A', B').$$

Let γ be a segment connecting A' and B'. Then

$$d(A', B') = \text{length}(\gamma) = \text{length}(p \circ \gamma).$$

Note that $p \circ \gamma$ is a curve between $\bar{A}$ and $\bar{B}$. Thus $p \circ \gamma$ is a minimal geodesic. This shows that $(\mathbb{R}^2/\sim, \bar{d})$ is a length metric space. □

Together with the exercise below, $(\mathbb{R}^2./\sim, \bar{d})$ is 2-dimensional locally Euclidean.

Exercise 2.1.17. *Show that the quotient map $p : B_r(A) \to B_r(p(A))$ is bijective for $r > 0$ small.*

We can establish corresponding for the real projective plane $(\mathbb{S}^2/\sim, \bar{d})$ as well.

Question 2.1.18. *Can we classify (have a list) of all 2-dimensional locally Euclidean / spherical spaces?*

2.2. Groups

Definition 2.2.1. Let G be a set. A multiplication of G is a map (operation) $\cdot : G \times G \to G$, which we usually write $\cdot(a, b)$ as $a \cdot b$, satisfying
(1) there is an element $e \in G$ such that $e \cdot a = a \cdot e = a$ for any $a \in G$;
(2) for each $a \in G$, there is an element a^{-1} such that $a \cdot a^{-1} = a^{-1} \cdot a = e$;
(3) $(a \cdot b) \cdot c = a \cdot (b \cdot c)$ for all $a, b, c \in G$.
We call e the *identity element*, a^{-1} the *inverse* of a, and $(G, \cdot)$ a *group*.

Examples 2.2.2.
• $(\mathbb{Z}, +)$. The identity element is 0. The inverse of $a \in \mathbb{Z}$ is $-a$.
• $(\mathbb{R}_+, \cdot)$, where $\cdot$ is the usual multiplication. The identity element is 1. The inverse of $a \in \mathbb{R}_+$ is $1/a$.
• $(\mathbb{Z}_p, +)$, the so called modulus that you learned in Math 8. $\mathbb{Z}_p = \{\bar{0}, \bar{1}, ..., \overline{p-1}\}$. $+$ is defined as follows: for $\bar{a}, \bar{b} \in \mathbb{Z}_p$, there is a unique element $c \in \{0, 1, ..., p-1\}$ such that $a + b \equiv c(\text{mod } p)$, then we define $\bar{a} + \bar{b}$ as $\bar{c}$. For instance, in $\mathbb{Z}_4$, we have $\bar{2} + \bar{2} = \bar{0}$, $\bar{2} + \bar{3} = \bar{1}$.
• The set of all translations in $\mathbb{R}^2$ along $x-$axis by a unit $2k\pi$, where $k \in \mathbb{Z}$. The composition of motions gives the multiplication.
• Let ϕ be the rotation around the origin by 90 degrees in $\mathbb{R}^2$. Denote id as the identity map of $\mathbb{R}^2$. Then $\{\text{id}, \phi, \phi^2, \phi^3\}$ forms a group with multiplication as the composition. Note that $\phi^4 = \text{id}$.
• Permutation group S_n is the set of all bijections from $\{1, 2, ..., n\}$ to itself.
• The circle group $(S^1, \cdot)$. We treat S^1 as the unit circle in the complex plane. In this way, we can write each element as $e^{i\theta}$. $\cdot$ is the multiplication of complex numbers, that is, $e^{i\theta_1} \cdot e^{i\theta_2} = e^{i(\theta_1+\theta_2)}$.
• General linear group $GL_n(\mathbb{R})$: the set of all $n \times n$ real matrices with non-zero determinant. The multiplication is the usual multiplication of two matrices.
• Special linear group $SL_n(\mathbb{R})$: the set of all $n \times n$ real matrices A with $\det(A) = 1$. This is a group: if $\det(A) = \det(B) = 1$, then $\det(AB) = \det(A)\det(B) = 1$.
• Orthogonal group $O(n)$: the set of all $n \times n$ orthogonal matrices A (, that is, $AA^T = I_n$, where A^T is the transport of A and I_n is the $n \times n$ identity matrix). As an exercise, have a try to prove that $O(n)$ is indeed a group.
• Special linear group $SO(n) = O(n) \cap SL_n(\mathbb{R})$.

Definition 2.2.3. Let S be a subset of a group G. We say that S generated G, if any element of G can be written as a product of elements in S or S^{-1}, where $S^{-1} = \{a^{-1} | a \in S\}$.

Examples 2.2.4.
• $(\mathbb{Z}, +)$ is generated by a single element $\{1\}$.
• $(\mathbb{Z}_p, +)$ is generated by $\{\bar{1}\}$.
• In the fourth example of Examples 2.2, the translation by 2π generates the group.
• In the fifth example of Examples 2.2, ϕ generates the group.
• S_n can be generated by transpositions.
• The circle group can be generated by a small neighborhood around e.
• As a (nontrivial) fact, the set of reflections generates $O(n)$.

Definition 2.2.5. Let $(G, \cdot)$ be a group. We say that a subset H of G is a *subgroup* if $(H, \cdot)$ is a group, or equivalently, H is closed under multiplication and taking inverses. This is usually denoted as $H \leq G$.

Examples 2.2.6.
• $(\mathbb{Z}, +) \leq (\mathbb{Q}, +) \leq (\mathbb{R}, +)$.
• $(S^1, \cdot) \leq (\mathbb{C} - \{0\}, \cdot)$.
• $SO(n) \leq SL_n(\mathbb{R}) \leq GL_n(\mathbb{R})$.

Definition 2.2.7. Let G and H be two groups. A *group homomorphism* is a map $\psi : G \to H$ such that
$$\psi(g_1 \cdot g_2) = \psi(g_1) \cdot \psi(g_2)$$

for all $g_1, g_2 \in G$.

We say that a group homomorphism $\psi : G \to H$ is a *group isomorphism* if it is bijective.

Examples 2.2.8.
• The determinant $\det : GL_n(\mathbb{R}) \to (\mathbb{R} - \{0\}, \cdot)$ is a group homomorphism because

$$\det(AB) = \det(A)\det(B).$$

• In the fourth example of Examples 2.2.2, the group is isomorphic to $\mathbb{Z}$. The isomorphism can be realized by sending the translation by 2π to $1 \in \mathbb{Z}$.
• In the fifth example of Examples 2.2.2, the group is isomorphic to $\mathbb{Z}_4$. The isomorphism sends ϕ to $\{\bar{1}\} \in \mathbb{Z}_4$.

Proposition 2.2.9. *Let (X, d) be a metric space. Then the set of all isometries of (X, d) forms a group with multiplication as composition of maps. This group is denoted as* $\mathrm{Isom}(X)$.

PROOF. The identity element is the identity map of X. We show that $\mathrm{Isom}(X)$ is closed under taking inverses and multiplications.

Let $F \in \mathrm{Isom}(X)$. Since F is bijective, it has an inverse F^{-1}. Clearly F^{-1} is also a bijection. We show that F^{-1} preserves the metric:

$$d(F^{-1}(x), F^{-1}(y)) = d(F(F^{-1}(x)), F(F^{-1}(y))) = d(x, y).$$

Let $F_1, F_2 \in \mathrm{Isom}(X)$. We show that $F_1 \circ F_2 \in \mathrm{Isom}(X)$:

$$d((F_1 \circ F_2)(x), (F_1 \circ F_2)(y)) = d(F_2(x), F_2(y)) = d(x, y).$$

□

Remark 2.2.10. If two spaces are isometric, then their isometry groups are isomorphic. In principle, this is true because isometry group comes from the metric, since the metrics are the same, then the isometry groups should be the same as well. To prove this formally, let $F : (X, d_X) \to (Y, d_Y)$ be an isometry. For $g \in \mathrm{Isom}(X)$, its conjugation under F, that is, $F \circ g \circ F^{-1}$, is an isometry of (Y, d_Y). This provides a map

$$C_F : \mathrm{Isom}(X) \to \mathrm{Isom}(Y), \quad g \mapsto F \circ g \circ F^{-1}.$$

Exercise 2.2.11. *Prove that C_F is indeed a group isomorphism.*

2.3. Group actions on Geometries

Definition 2.3.1. Let X be a set and G be a group. A (left) *G-action on X* is a map $\mu : G \times X \to X$ satisfying
(1) $\mu(e, x) = x$ for any $x \in X$;
(2) $\mu(gh, x) = \mu(g, \mu(h, x))$ for all $g, h \in G$ and $x \in X$.
When the map μ is clear, we usually write $\mu(g, x)$ as $g(x)$ or $g \cdot x$. With this notation, condition (2) can be written as $(gh) \cdot x = g \cdot (h \cdot x)$.

Examples 2.3.2.
• The permutation group S_n acts on the set $\{1, 2, ..., n\}$ by, well, permutations.
• $\mathbb{Z}$ acts on $\mathbb{R}^2$ by translations: $\mu(k, (x, y)) = (x + 2k\pi, y)$.
• $\mathbb{Z}$ acts on $\mathbb{R}^2$ by glide reflections: $\mu(k, (x, y)) = (x + 2k\pi, (-1)^k y)$.
• S^1 acts on $\mathbb{R}^2$ by rotations around the origin:

$$\mu(\theta, (x, y)) = (x \cos\theta + y \sin\theta, -x \sin\theta + y \cos\theta).$$

• A group acts on itself by left translations: $\mu(g,h) = g \cdot h$.
• A group acts on itself by conjugations: $\mu(g,h) = ghg^{-1}$. As an exercise, prove this is a group action.

Exercise 2.3.3. *Let $\mathbb{H}$ be the upper half complex plane $\mathbb{H} = \{z \in \mathbb{C} \mid \mathrm{Im}(z) > 0\}$. Let $SL_2(\mathbb{R})$ be the special linear group of $\mathbb{R}^2$. We define a map*

$$\mu : SL_2(\mathbb{R}) \times \mathbb{H} \to \mathbb{C}, \quad \left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}, z \right) \mapsto \frac{az+b}{cz+d}.$$

Show that
(1) μ has image in $\mathbb{H}$.
(2) μ defines a $SL_2(\mathbb{R})$-action on $\mathbb{H}$.

Definition 2.3.4. Let X be a G-space and let $x \in X$. The $G-$*orbit* at x is defined as

$$G \cdot x = \{g \cdot x | g \in G\}.$$

As an exercise, think about what is the G-orbit in Examples 2.3.2.

We define an equivalence relation on X: $x_1 \sim x_2$, if x_1 and x_2 belong to the same $G-$orbit. In other words, there is $g \in G$ such that $g \cdot x_1 = x_2$.

Proposition 2.3.5. *$\sim$ in an equivalence relation on X.*

Proof. Transitivity is the only non-trivial part. Suppose that $x_1 \sim x_2$ and $x_2 \sim x_3$. By definition, this means that there is g_1 and g_2 in G such that

$$x_2 = g_1 \cdot x_1, \quad x_3 = g_2 \cdot x_2.$$

Then

$$x_3 = g_2 \cdot x_2 = g_2 \cdot (g_1 \cdot x_1) = (g_2 g_1) \cdot x_1.$$

□

The quotient space with respect the above equivalence relation is called the *orbit space*, and is denoted as X/G. We usually write $\bar{x}$ as an element in X/G.

Definition 2.3.6. Let (X, d) be a metric space. We say a G-action on X is isometric (or G acts on X by isometries), if

$$\mathcal{A}_g : X \to X, \quad x \to g \cdot x$$

is an isometry for each $g \in G$.

Let (X, d) be a length metric space with an isometric G-action. We want to define the quotient metric on X/G to be minimal between two orbits, that is,

$$\bar{d}(\bar{x}, \bar{y}) = \min_{g_1, g_2 \in G} d(g_1 \cdot x, g_2 \cdot y)$$

Since G acts by isometries, RHS equals to

$$\min_{g_1, g_2 \in G} d(x, g_1^{-1} g_2 \cdot y) = \min_{h \in G} d(x, h \cdot y).$$

Ideally, we wish that this minimal can be achieved. However, this is not always true. For example, consider $\mathbb{Q}$-action by $(\mathbb{R}, d_0)$ by translations: $\mu(q,x) = q + x$. Since the orbit $\mathbb{Q} \cdot 0$ is dense in $\mathbb{R}$, for any $x \in \mathbb{R}$, the minimal (infimum) of $d(q \cdot 0, x)$, which is 0, cannot be attained.

Motivated by the cylinder example, we define

Definition 2.3.7. Let (X, d) be a metric space. We say that a subset $S \subset X$ is discrete in X, if for any $s \in S$, there is $r > 0$ such that $B_r(s) \cap S = \{s\}$.

Let G be an isometric action on (X, d). We say that G-action is discrete, if the orbit $G \cdot x$ is discrete for all $x \in X$.

Example 2.3.8. The set $\{1/n | n \in \mathbb{Z}\}$ is not discrete in $(\mathbb{R}, d_0)$.

For an orbit $G \cdot x$ to be discrete, we can indeed use a weaker condition.

Lemma 2.3.9. *Let (X, d_X) be a G-space and let $x \in X$. If there is $s \in G \cdot x$ and $r > 0$ such that $B_r(s) \cap (G \cdot x) = \{s\}$, then $G \cdot x$ is discrete.*

Proof. By assumption, there is $s \in G \cdot x$ and $r > 0$ such that $B_r(s) \cap (G \cdot x) = \{s\}$. We show that this $r > 0$ also works for other points in the orbit, that is, $B_r(g \cdot s) \cap (G \cdot x) = \{g \cdot s\}$ for any $g \in G$.

Suppose the contrary, then there are a distinct orbit point $h \cdot s \neq g \cdot s$ lying in $B_r(g \cdot s)$. We consider the point $(g^{-1}h) \cdot s$. This point is not s; otherwise $(g^{-1}h) \cdot x$ would give $h \cdot x = g \cdot x$. Also,

$$d((g^{-1}h) \cdot s, s) = d(h \cdot s, g \cdot s) < r,$$

thus $(g^{-1}h) \cdot s \in B_r(s)$. This contradicts our choice of r. □

Other than requiring G-action to be discrete, we need another reasonable condition to make sure the minimum can be obtained. The Definition 2.3.10 and Proposition 2.3.11 below are not required for this course.

Definition 2.3.10. We say that a metric space (X, d) is locally compact, if for each $x \in X$, there is $r > 0$ such that the local ball $B_r(x)$ satisfies that any sequence in $B_r(x)$ has a convergent subsequence with a limit in X.

Proposition 2.3.11. *Let (X, d) be a locally compact metric space with a discrete isometric G-action. For two non-equivalent points x and y, $min_{h \in G} d(x, h \cdot y)$ can be achieved and is positive.*

Proof. We can always take the infimum. Let $m = \inf_{h \in G} d(x, h \cdot y)$. There is a sequence of point $p_i = h_i \cdot y$ such that $d(x, p_i) \to m$ as $i \to \infty$. The tail of this sequence lies in $B_{2m}(x)$. By assumption, $\{p_i\}$ has a convergent subsequence. This contradicts to the fact that $G \cdot y$ is discrete. □

For spaces like $\mathbb{R}^n$, which is locally compact, there is a easier way to prove.

Lemma 2.3.12. *Let G be an isometric action on $(\mathbb{R}^n, d_0)$. Suppose that G-action is discrete, then for any $x, y \in X$ and any $R > 0$, $B_R(x) \cap (G \cdot y)$ is a finite set.*

Proof. Let $r > 0$ such that points of $B_r(y) \cap (G \cdot y) = \{y\}$. By Lemma 2.3.9, $B_r(s) \cap (G \cdot y) = \{s\}$ for any $s \in G \cdot y$. We claim that

$$B_{r/2}(s_1) \cap B_{r/2}(s_2) = \emptyset,$$

for all $s_1 \neq s_2 \in G \cdot y$. In fact, suppose that $z \in B_{r/2}(s_1) \cap B_{r/2}(s_2)$, then by triangle inequality, we have

$$d(s_1, s_2) \leq d(s_1, z) + d(z, s_2) < r/2 + r/2 = r.$$

Thus $s_2 \in B_r(s_1) \cap (G \cdot y)$, a contradiction. This verifies the claim.

Note that if $s \in B_R(x)$, then $B_{r/2}(s) \subset B_{R+r/2}(x)$ by triangle inequality.Thus family of balls $\{B_{r/2}(s)\}_{s\in B_R(x)\cap(G\cdot y)}$ are all contained in $B_{R+r/2}(x)$ and pairwise disjoint. We have

$$\mathrm{vol}(B_{R+r/2})(x) \geq \sum_{s\in B_R(x)\cap(G\cdot y)} \mathrm{vol}(B_{r/2}(s)) = |B_R(x)\cap(G\cdot y)| \cdot \mathrm{vol}(B_{r/2}(0)).$$

Thus $B_R(x)\cap(G\cdot y)$ has cardinality

$$|B_R(x)\cap(G\cdot y)| \leq \frac{\mathrm{vol}(B_{R+r/2}(0))}{\mathrm{vol}(B_{r/2}(0))} = \frac{(R+r/2)^n}{(r/2)^n} < \infty.$$

□

A similar argument works for spaces like the sphere S^n.

Now we know that $B_R(x)\cap(G\cdot y)$ is a finite set. As a result, $\min_{h\in G} d(x, h\cdot y)$ can be attained and is positive. With this, we can show that $\bar{d}$ on X/G is a distance function.

Proposition 2.3.13. *(1) $\bar{d}$ on X/G is a metric.*
(2) If in addition (X,d) is a length metric, then so is $(X/G,\bar{d})$.

Proof. (1) $\bar{d}(\bar{x},\bar{y}) = \min_{h\in G} d(x, h\cdot y) > 0$ guarantees than the first condition in Definition 1.1 is fulfilled. The second condition clearly holds. We check the triangle inequality, which is indeed similar to the cylinder case.

For $x, y, z \in X$, we choose $z_0 \in G\cdot z$ such that $\bar{d}(\bar{x},\bar{z}) = d(x, z_0)$. Then we choose $y_0 \in G\cdot y$ such that $\bar{d}(\bar{z},\bar{y}) = d(z_0, y_0)$. We have

$$\bar{d}(\bar{x},\bar{y}) \leq d(x,y_0) \leq d(x,z_0) + d(z_0,y_0) = \bar{d}(\bar{x},\bar{z}) + \bar{d}(\bar{z},\bar{y}).$$

We need a lemma before proving (2). □

Lemma 2.3.14. *Let $p : (X,d) \to (X/G,\bar{d}), x \mapsto \bar{x}$ be the quotient map. Let $\gamma : I \to X$ be a path of finite length in X. Then* $\mathrm{Length}(p\circ\gamma) \leq \mathrm{Length}(\gamma)$.

Proof. This follows from the definition of $\bar{d}$ that p is distance non-increasing: $\bar{d}(\bar{x},\bar{y}) \leq d(x,y)$ for all $x, y \in X$. Let $\pi = \{t_0, ..., t_k, ...t_n\}$ be any partition of I. Then

$$\begin{aligned}
\mathrm{Length}(p\circ\gamma) &= \lim_{||\pi||\to 0} \sum_{i=0}^{n-1} \bar{d}(p\circ\gamma(t_i), p\circ\gamma(t_{i+1})) \\
&\leq \lim_{||\pi||\to 0} \sum_{i=0}^{n-1} d(\gamma(t_i), \gamma(t_{i+1})) \\
&= \mathrm{Length}(\gamma).
\end{aligned}$$

□

Proof of Proposition 2.3.13(2). Given $\bar{x},\bar{y} \in X/G$, we need to find a path between them whose length equals to $\bar{d}(\bar{x},\bar{y})$. Let $x, y \in X$ such that $d(x,y) = \bar{d}(\bar{x},\bar{y})$. Because (X,d) is a length metric space, there is a path γ from x to y with $\mathrm{Length}(\gamma) = d(x,y)$. By the previous lemma,

$$\mathrm{Length}(p\circ\gamma) \leq \mathrm{Length}(\gamma) = d(x,y) = \bar{d}(\bar{x},\bar{y}).$$

On the other hand, because $p\circ\gamma$ connects $\bar{x}$ and $\bar{y}$, we have $\mathrm{Length}(p\circ\gamma) \geq \bar{d}(\bar{x},\bar{y})$ by triangle inequality. Hence $\mathrm{Length}(p\circ\gamma) = \bar{d}(\bar{x},\bar{y})$. □

Definition 2.3.15. We say that G-action is free, if for any $g \neq e$ and any $x \in X$, $g \cdot x \neq x$ holds.

In other words, a free G-action means any element $g \in G - \{e\}$ does not have fixed points on X.

Lemma 2.3.16. *Let G be a subgroup of* $\mathrm{Isom}(\mathbb{R}^n)$. *If G-action on $\mathbb{R}^n$ is free and discrete, then there is $D > 0$ such that $d(g \cdot x, x) \geq D$ for all $g \neq e \in G$ and all $x \in X$.*

We will prove this lemma later for $\mathbb{R}^2$ when we have a better understanding on isometries of $\mathbb{R}^2$. Assuming this at the moment, we prove

Theorem 2.3.17. *Let G be a subgroup of* $\mathrm{Isom}(\mathbb{R}^n)$. *If G-action on $\mathbb{R}^n$ is free and discrete, then the quotient space $(X/G, \bar{d})$ is n-dimensional locally Euclidean.*

Proof. Let $D > 0$ be the constant in Lemma 2.3.16. We claim that if $d(x, y) < D/2$, then $\bar{d}(\bar{x}, \bar{y}) = d(x, y)$. In fact, for any $x' \in G \cdot x$ that is not x, by triangle inequality, we have

$$d(x', y) \geq d(x', x) - d(x, y) \geq D - D/2 = D/2 > d(x, y).$$

This shows that

$$d(x, y) = \min_{x' \in G \cdot x} d(x', y) = \bar{d}(\bar{x}, \bar{y}).$$

For any $y, z \in B_{D/4}(x)$, by triangle inequality

$$d(y, z) \leq d(y, x) + d(x, z) < D/4 + D/4 = D/2.$$

It follows from the claim that $\bar{d}(\bar{y}, \bar{z}) = d(y, z)$. This shows that the quotient map $p : X \to X/G$, when restricting it to $B_{D/4}(x)$, preserves the metric.

Next we show that $p|_{B_{D/4}(x)}$ is a bijection between $B_{D/4}(x)$ and $B_{D/4}(\bar{x})$; this would finish the proof that $B_{D/4}(\bar{x})$ is isometric to $D/4$-ball in $\mathbb{R}^n$.

$p|_{B_{D/4}(x)}$ is injective: Suppose that $y_1, y_2 \in B_{D/4}(x)$ maps to the same point through p. This means that y_1 and y_2 lie in the same orbit with $d(y_1, y_2) < D/2$, which contradicts our choice of D.

$p|_{B_{D/4}(x)}$ is onto $B_{D/4}(\bar{x})$: For any $\bar{y}$ with $\bar{d}(\bar{x}, \bar{y}) < D/4$. Recall that $\bar{d}(\bar{x}, \bar{y})$ is the minimal between x and any point equivalent to y. Therefore, there is a point y' such that $p(y') = \bar{y}$ and $d(x, y') = \bar{d}(\bar{x}, \bar{y}) < D/4$, that is, $y' \in B_{D/4}(x)$. □

CHAPTER 3

Free and discrete isometric actions on $\mathbb{R}^2$ or $\mathbb{S}^2$

3.1. Hello again, linear algebra

In this section, we make use of linear algebra to prove the following:

Theorem 3.1.1. *Every element g of $\text{Isom}(\mathbb{R}^n)$ can be uniquely expressed in the form of $g(x) = Ax + v$, where $A \in O(n)$ and $v \in \mathbb{R}^n$.*

Lemma 3.1.2. *Every element g of $\text{Isom}(\mathbb{R}^n)$ can be uniquely written as a composition $T \circ h$, where T is a translation and h is an isometry fixing the origin.*

Proof. Let $v = g(0) \in \mathbb{R}^n$. If $v = 0$, then g itself is an isometry fixing the origin and the statement holds clearly. We assume that $v \neq 0$. We define isometries

$$T : \mathbb{R}^n \to \mathbb{R}^n, \quad x \mapsto x + v;$$

$$h : \mathbb{R}^n \to \mathbb{R}^n, \quad x \mapsto g(x) - v.$$

h is an isometry fixing the origin because it is the composition of two isometries and

$$h(0) = g(0) - v = v - v = 0.$$

Note that

$$T \circ h(x) = T(g(x) - v) = g(x) - v + v = g(x).$$

The uniqueness part is left as an exercise. □

Lemma 3.1.3. *Let $h : \mathbb{R}^n \to \mathbb{R}^n$ be a function. The followings are equivalent.*
(1) h is an isometry fixing the origin.
(2) h preserves the dot product: $h(x) \cdot h(y) = x \cdot y$ for all $x, y \in \mathbb{R}^n$.
(3) h is linear and its matrix representation is orthogonal.

Proof. First recall that the dot product and the distance is related by

$$d(x, y)^2 = (x - y) \cdot (x - y).$$

We write $x \cdot x = ||x||^2$ below for simplicity.

(1)⇒(2): Suppose that h is an isometry. This means that

$$||h(x) - h(y)|| = d(h(x), h(y)) = d(x, y) = ||x - y||$$

for any $x, y \in \mathbb{R}^n$. Squaring both sides, we have

$$(h(x) - h(y)) \cdot (h(x) - h(y)) = (x - y) \cdot (x - y)$$
$$h(x) \cdot h(x) - 2h(x) \cdot h(y) + h(y) \cdot h(y) = x \cdot x - 2x \cdot y + y \cdot y$$
$$||h(x)||^2 - 2h(x) \cdot h(y) + ||h(y)||^2 = ||x||^2 - 2x \cdot y + ||y||^2.$$

Together with

$$||h(x)|| = d(h(x), 0) = d(h(x), h(0)) = d(x, 0) = ||x||,$$

it follows that

$$h(x) \cdot h(y) = x \cdot y.$$

We also need to show that $h(0) = 0$:

$$||h(0)||^2 = h(0) \cdot h(0) = 0 \cdot 0 = 0.$$

(2)⇒(1): Suppose that h preserves the dot product. Then for any $x, y \in \mathbb{R}^n$,

$$\begin{aligned} d(h(x), h(y))^2 &= (h(x) - h(y)) \cdot (h(x) - h(y)) \\ &= h(x) \cdot h(x) - 2h(x) \cdot h(y) + h(y) \cdot h(y) \\ &= x \cdot x - 2x \cdot y + y \cdot y \\ &= (x - y) \cdot (x - y) \\ &= d(x, y)^2 \end{aligned}$$

(2)⇒(3): We first show that h is linear. This means that we need to show that

$$h(ax + by) = ah(x) + bh(y)$$

for any $x, y \in \mathbb{R}^n$ and any $a, b \in \mathbb{R}$. Let $\{e_1, e_2, ..., e_n\}$ be a standard basis of $\mathbb{R}^n$. Since h preserves the dot product, $\{h(e_1), ..., h(e_n)\}$ is an orthonormal basis of $\mathbb{R}^n$. For each $j = 1, ..., n$, we have

$$\begin{aligned} h(ax + by) \cdot h(e_j) &= (ax + by) \cdot e_j = a(x \cdot e_j) + b(y \cdot e_j) \\ &= a(h(x) \cdot h(e_j)) + b(h(y) \cdot h(e_j)) \\ &= (ah(x) + bh(y)) \cdot h(e_j). \end{aligned}$$

Then $h(ax + by) = ah(x) + bh(y)$ follows from the fact that two vector in $\mathbb{R}^n$ are equal if and only if they have the same dot product with each element in an orthonormal basis.

Next we show that the matrix representation of h (under the standard basis) is orthogonal. Let A be its matrix representation, that is, $h(x) = Ax$, where Ax is understood as matrix multiplication. For all $x, y \in \mathbb{R}^n$ as column vectors (under the standard basis), we have

$$x \cdot y = h(x) \cdot h(y) = Ax \cdot Ay = (Ax)^T (Ay) = x^T (A^T A) y.$$

For $x = e_i$ and $y = e_j$, the term $x^T (A^T A) y$ is the (i, j)-th entry of the matrix $A^T A$. Hence the (i, j)-th entry of $A^T A$ should be $e_i \cdot e_j = \delta_{i,j}$. This shows that $A^T A$ is the identity matrix, that is, A is orthogonal.

(3)⇒(2): Let $x, y \in \mathbb{R}^n$. We regard x, y as column vectors under the standard basis. Then the dot product $x \cdot y$ equals multiplication $x^T y$. Since $h : x \mapsto Ax$, where A is orthogonal, we have

$$h(x) \cdot h(y) = Ax \cdot Ay = (Ax)^T (Ay) = x^T (A^T A) y = x^T y = x \cdot y.$$

□

Theorem 3.1.1 follows directly from Lemmas 3.1.2 and 3.1.3.

Proof of Theorem 3.1.1. Let $g \in \text{Isom}(\mathbb{R}^n)$. By Lemma 3.1.2, we know that g can be written as

$$g : x \mapsto h(x) + v,$$

where h is an isometry fixing the origin and $v \in \mathbb{R}^n$. Applying Lemma 3.1.3, h is linear and its matrix representation A is orthogonal. Thus

$$g(x) = h(x) + v = Ax + v.$$

□

From Theorem 3.1.1, we know that every element g of Isom($\mathbb{R}^n$) is in the form of $g(x) = Ax + v$, where $A \in O(n)$ and $v \in \mathbb{R}^n$. We write this element g as (A, v). By calculating the composition of (A, v) and (B, w), it is easy to get a formula for group multiplication:

$$\begin{aligned}(A, v) \cdot (B, w)(x) &= (A, v)(Bx + w) = A(Bx + w) + v \\ &= ABx + Aw + v = (AB, Aw + v)(x).\end{aligned}$$

Thus

$$(A, v) \cdot (B, w) = (AB, Aw + v).$$

Using this rule for multiplication, we can also find the inverse of (A, v):

$$(A, v)^{-1} = (A^{-1}, -A^{-1}v).$$

In fact, we can check that

$$(A, v) \cdot (A^{-1}, -A^{-1}v) = (AA^{-1}, A(-A^{-1}v) + v) = (I, 0),$$
$$(A^{-1}, -A^{-1}v) \cdot (A, v) = (A^{-1}A, A^{-1}v + (-A^{-1}v)) = (I, 0).$$

3.2. Elements of Isom($\mathbb{R}^2$) without fixed points

Now we focus on $\mathbb{R}^2$. Recall that we are looking for isometries that does not have fixed points.

Proposition 3.2.1. *If $g \in$ Isom($\mathbb{R}^2$) does not have any fixed points, then g is a translation or a glide reflection.*

Proof. We write $g = (A, v)$. Since it has no fixed points, the equation $Ax+v = x$ has no solutions, that is, $(A - I)x = -v$ has no solution. By basic linear algebra, we conclude $\det(A - I) = 0$. In other words, A has an eigenvalue 1. Recall that $A \in O(2)$ and thus $\det(A) = \pm 1$. Hence the product of two eigenvalues should be 1 or -1. There are only two cases:
Case 1: The other eigenvalue is also 1. Because $A \in O(2)$, A must be the identity matrix in this case (this step is left as an exercise). Then $g = (I, v)$ is a translation.
Case 2: The other eigenvalue is -1. Let e_1 and e_2 be the unit length eigenvector with eigenvalue 1 and -1, respectively. By linear algebra, these two vectors are orthogonal. Under the orthonormal basis $\{e_1, e_2\}$, we can rewrite $g = (F, w)$, where F is the diagonal matrix diag$\{1, -1\}$. F is indeed the reflection about the line $\mathbb{R}e_1$. With this, the fact that $(F - I)x = w$ has no solutions translates to that

$$\begin{pmatrix} 0 & 0 \\ 0 & -2 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} w_1 \\ w_2 \end{pmatrix}$$

has no solutions. We conclude that $w_1 \neq 0$. Note that

$$g(x) = Fx + w = Fx + w_1e_1 + w_2e_2 = (Fx + w_2e_2) + w_1e_1 = h(x) + w_1e_1,$$

where $h(x) = Fx + w_2e_2$ Because F reflects about $\mathbb{R}e_1$ and e_2 is perpendicular to e_1, the map h is indeed the reflection about the line $L = \{y = w_2/2\}$. The translation left $w_1e_1 \neq 0$ is parallel to L. We move the coordinate up (along e_2) by

$w_2/2$. Then under the new coordinate, we can write g as (F, w_1e_1) with $w_1 \neq 0$, which is a glide reflection. □

Remark 3.2.2. With the proof above, we also showed that any glide reflection of $\mathbb{R}^2$ can be written as (F, w_1e_1) under a suitable coordinate $\{e_1, e_2\}$.

3.3. Subgroups of Isom($\mathbb{R}^2$)

With Proposition 3.2.1, we proceed to classify subgroups of Isom($\mathbb{R}^2$), whose actions are free and discrete. Let Γ be such a subgroup. By Proposition 3.2.1, we know that any non-identity element of Γ is either a translation or a glide reflection. We define T be the set of all translations in Γ plus the identity element. It is clear that T is a subgroup of Γ. We classify the types of Γ by how many directions the translations can move:

- Type I. $T = \{e\}$;
- Type II. T consists of translations in one direction;
- Type III. T contains at least two translations along two linearly independent vectors;

Depending on whether Γ contains glide reflections, we further classify type II and III, respectively, into:

- Type IIa and IIIa: $\Gamma = T$, that is, Γ does not contain glide reflections;
- Type IIb and IIIb: Γ contains glide reflections.

For Type I above, $T = \{e\}$ implies $\Gamma = \{e\}$. In fact, suppose that Γ contains a glide reflection $g = (F, w_1e_1)$, then

$$g^2 = (F^2, F(w_1e_1) + w_1e_1) = (I, 2w_1e_1)$$

would be a translation in Γ, which contradicts to $T = \{e\}$.

The following Lemma will be used when we study Type IIb and Type IIIb.

Lemma 3.3.1. *Suppose that Γ contains a glide reflection g, then $\Gamma = T \cup Tg$, where $Tg = \{t \cdot g \mid t \in T\}$.*

Proof. By Remark 3.2.2, we can write $g = (F, w)$ under a suitable coordinate, where w is a multiple of e_1. Then $g^{-1} = (F^{-1}, -F^{-1}w) = (F, -w)$. For any other glide reflection $g' = (A, v) \in \Gamma$, we show that the element $g'g^{-1} \in \Gamma$ is a translation or identity. In fact,

$$g'g^{-1} = (A, v) \cdot (F, -w) \cdot (A, a) = (AF, -Aw + v).$$

Because A has determinant -1 (by the proof of Proposition 3.2.1), AF has determinant 1. When $g'g^{-1}$ is not identity, $(AF, -Aw + v)$ is a non-trivial element in Γ; in particular, this element does not have fixed points. As a result, AF must be the identity matrix (again by the proof of Proposition 3.2.1). Thus $g^{-1}g' = (I, -Aw+v)$ is a translation we denote as t. This shows $g' = t \cdot g$. In other words, any glide reflection of Γ is an element of Tg. Because Γ only contains translations and glide reflections, we see that $\Gamma = T \cup Tg$. □

3.3.1. **Type II.** Recall that type II means that T, all translations of Γ, are in the same direction. Let e_1 be the unit vector parallel to this direction, then any translation of Γ is in form of (I, ce_1) for some $c \in \mathbb{R}$. For type IIa, we have $\Gamma = T$. Using the fact that Γ-action is discrete, we prove the following:

Proposition 3.3.2. *Suppose that Γ is of type IIa. Then under a suitable orthonormal coordinate $\{e_1, e_2\}$, there is $a > 0$ such that (I, ae_1) generates Γ.*

In general, if a set S generates a group Γ, we write $\Gamma = \langle S \rangle$. With this notation, Proposition 3.3.2 says

$$\Gamma = \langle (I, ae_1) \rangle = \{(I, kae_1) | k \in \mathbb{Z}\}.$$

PROOF. We know that any element of Γ are in the form of (I, ce_1) for some $c \in \mathbb{R}$. We choose $a > 0$ so that (I, ae_1) is the shortest translation in Γ. In other words,

$$a = \min\{|c| : (I, ce_1) \in \Gamma - \{e\}\}.$$

The minimum exists because there are only finitely many points in $B_r(0)$ that are orbit points of 0.

We claim that any element of Γ is a multiple of (I, ae_1). Suppose that $(I, be_1) \in \Gamma$ such that b/a is not an integer. We write $b/a = k+r$, where $k \in \mathbb{Z}$ and $r \in (0, 1)$. Since $(I, -kae_1) \in \Gamma$ and Γ is a group, we conclude that

$$(I, -kae_1) \cdot (I, be_1) = (I, rae_1) \in \Gamma.$$

This contradicts to the our choice of (I, ae_1) as the minimal translation. □

Proposition 3.3.3. *Suppose that Γ is of type IIb. Then under a suitable orthonormal coordinate $\{e_1, e_2\}$, there is $a > 0$ such that*

$$T = \langle (I, ae_1) \rangle, \quad \Gamma = \langle (F, ae_1/2) \rangle.$$

PROOF. Under a suitable coordinate, we can write g as (F, be_1). Note that $g^2 = (I, 2be_1)$ is a translation. Since all translation of Γ are in the same direction, by Proposition 3.3.2, there is $a > 0$such that the subgroup T can be generated by a single element (I, ae_1). Thus there is an integer k such that $ka = 2b$.

We claim that k is odd. Suppose that $k = 2l$ for some $l \in \mathbb{Z}$. Then $la = b$. Note that $(I, -lae_1)$ is an element of Γ, thus

$$(I, -lae_1) \cdot g = (I, -lae_1) \cdot (F, lae_1) = (F, 0) \in \Gamma.$$

We reached a contradiction because Γ-action should be free. This verifies the claim.

Because k is odd, we write $k = 2l + 1$ for some $l \in \mathbb{Z}$. So $(2l + 1)a = 2b$, that is, $b = (l + 1/2)a$. We consider

$$h = (I, -lae_1) \cdot g = (I, -lae_1) \cdot (F, (l + 1/2)ae_1) = (F, ae_1/2),$$

which is a glide reflection in Γ. $h^2 = (I, ae_1)$ generates T. Also, Proposition 3.3.1 assures that $\Gamma = T \cup Th$. As a result, h generates Γ.

We check that Γ-action is free and discrete. It is clear that Γ-action is free because $\Gamma = T \cup Th$ consists of translations and glide reflections. It remains to show it is discrete. For any $x \in \mathbb{R}^2$ and any $(I, kae_1) \in T - \{e\}$, we estimate:

$$d(x, (I, kae_1)(x)) = ||kae_1|| = |ka| \geq |a|.$$

Denote $p_1 : \mathbb{R}^2 \to \mathbb{R}e_1$ as the projection map to the first component. Then for any $(F, ae_1/2 + kae_1) \in Tg$ and any $x \in \mathbb{R}^2$,

$$\begin{aligned} d(x, (F, ae_1/2 + kae_1)(x)) &= d(x, Fx + ae_1/2 + kae_1) \\ &\geq d(p_1(x), p_1(Fx) + p_1(ae_1/2 + kae_1)) \\ &= d(p_1(x), p_1(x) + ae_1/2 + kae_1) \\ &\geq d(0, ae_1/2 + kae_1) \\ &= |a/2 + ka| \geq |a|/2. \end{aligned}$$

□

3.3.2. **Type III.**

Proposition 3.3.4. *Suppose that* Γ *is of type IIIa. Then under a suitable orthonormal coordinate* $\{e_1, e_2\}$*, there are* $a, b > 0$ *and a unit vector* v *that is linearly independent of* e_1 *such that*

$$\Gamma = \langle (I, ae_1), (I, bv) \rangle = \{(I, kae_1 + lbv) | k, l \in \mathbb{Z}\}.$$

Proof of Proposition 3.3.4. We start with one translation in Γ. Changing the coordinate if necessary, this translation moves along e_1 direction. We consider a subgroup H of Γ, which consists of identity and all translations along e_1 direction. H is clearly a subgroup. By Proposition 3.3.2, H is generated by a single element (I, ae_1). We will to find a second element (I, bv) so that they together generate Γ.

Let $L = \mathbb{R}e_1$ be the e_1-axis. The orbit $H \cdot 0$ lies in L. Let $\overline{0, ae_1}$ be the segment from 0 to ae_1. We choose $g \in \Gamma - H$ such that it is the closest orbit point to $\overline{0, ae_1}$, that is,

$$d(g(0), \overline{0, ae_1}) = \min_{g' \in \Gamma - H} d(g'(0), \overline{0, ae_1}).$$

We claim that $g(0)$ is also the closest orbit point in $(\Gamma - H) \cdot 0$ to the entire line L:

$$d(g(0), L) = \min_{g' \in \Gamma - H} d(g'(0), L);$$

in particular, the minimal on the right hand side above exists. We argue this by contradiction. Suppose that there is $g' \in \Gamma - H$ such that $d(g'(0), L) < d(g(0), L)$. Let pe_1 be the point on L with $d(g'(0), L) = d(g'(0), pe_1)$. We write $p = ka + r$, where $k \in \mathbb{Z}$ and $r \in [0, a)$. We consider $g'' = (I, -kae_1) \cdot g' \in \Gamma - H$. Then

$$d(g'(0), L) = d(g'(0), p) = d((I, -kae_1) \cdot g'(0), (I, -kae_1) \cdot p) = d(g''(0), re_1).$$

Therefore, we find $g'' \in \Gamma - H$ such that

$$d(g''(0), \overline{0, ae_1}) = d(g''(0), re_1) = d(g'(0), L) < d(g(0), L) = d(g(0), \overline{0, ae_1}).$$

This contradicts to our choice of g. Hence the claim is verified.

As a translation, g can be written as (I, bv). (I, ae_1) and (I, bv) together generates a subgroup

$$\{(I, kae_1 + lbv) | k, l \in \mathbb{Z}\}.$$

Its orbit at 0 gives a lattice:

$$\mathcal{L} = \{kae_1 + lbv | k, l \in \mathbb{Z}\}.$$

Consider the lines that are parallel to e_1 or v and goes through $\mathcal{L}$. These lines divide the plane into many parallelogram of equal size. Suppose that Γ has an element α that not generated by (I, ae_1) and (I, bv). Then $\alpha(0)$ is not on the lattice $\mathcal{L}$. Note that $\alpha(0)$ cannot lie on horizontal line; otherwise, let $kae_1 + lbv$ be the orbit point that is left to $\alpha(0)$ on this line, then $(I, -kae_1 - lbv) \cdot \alpha(0)$ is a orbit point in L, whose distance to 0 is shorter than a, a contradiction to the choice of (I, ae_1). Now we pick a parallelogram containing $\alpha(0)$ and let $kae_1 + lbv$ be the left-bottom vertex of this parallelogram. Then $(I, -kae_1 - lbv) \cdot \alpha(0)$ is an orbit point in the parallelogram with left-bottom vertex 0. This orbit point has distance to L being shorter than $d(bv, L)$, which contradicts to our choice of $g = (I, bv)$. □

Proposition 3.3.5. *Suppose that Γ is of type IIIb. Then under a suitable orthonormal coordinate $\{e_1, e_2\}$, there are $a, b > 0$ such that*

$$T = \langle (I, ae_1), (I, be_2) \rangle, \quad \Gamma = \langle (F, \frac{1}{2}ae_1), (I, be_2) \rangle.$$

Proof. For any glide reflection in Γ, note that its square is a translation, which moves along a unit vector we call e_1. Let $(I, ae_1) \in \Gamma$ be the translation of minimal length along e_1. By the proof of Proposition 3.3.3, the glide reflection $g = (F, ae_1/2)$ must be an element of Γ. Let T be the subgroup of Γ consists of all translations. By the proof of Proposition 3.3.4, besides (I, ae_1), we can find another element (I, bv) so that these two elements together generates T.

We will show that $v = e_2$, the unit vector perpendicular to e_1. We write

$$bv = b_1 e_1 + b_2 e_2$$

for some $b_1, b_2 \in \mathbb{R}$. Compositing (I, bv) by some element (I, kae_1) if necessary, we can assume that $b_1 \in [0, a)$. We consider the element

$$\begin{aligned}(F, -ae_1/2) \cdot (I, bv) \cdot (F, ae_1/2) &= (F, ae_1/2) \cdot (F, ae_1/2 + bv) \\ &= (I, -ae_1/2 + F(ae_1/2 + bv)) \\ &= (I, F(bv)),\end{aligned}$$

which is a translation in Γ. Hence

$$(I, bv) \cdot (I, F(bv)) = (I, 2b_1 e_1) \in T.$$

Recall that $b_1 \in [0, a)$ and any translation along e_1 is in the form of (I, kae_1) for some $k \in \mathbb{Z}$. As a result, we conclude that $b_1 = 0$ or $b_1 = a/2$. We need to rule out the case $b_1 = a/2$. If $(I, ae_1/2 + b_2 e_2) \in \Gamma$, then

$$(I, -ae_1/2 - b_2 e_2) \cdot (F, ae_1/2) = (F, -b_2 e_2) \in \Gamma.$$

Observe that $(F, -b_2 e_2)$ is a reflection, which has fixed points; this contradicts to the assumption that Γ-action is free.

We summarize what we got so far: $b_1 = 0$, $g = (F, ae_1/2) \in \Gamma$, T is generated by two elements: (I, ae_1) and (I, be_2). By Lemma 3.3.1, $\Gamma = T \cup Tg$.

Finally, we check that Γ-action is free and discrete (this step confirms the existence of a group of type IIIb). It is clear that Γ-action is free because $\Gamma = T \cup Tg$ only consists of translations and glide reflections. It remains to show it is discrete. For any $x \in \mathbb{R}^2$ and any $(I, v) \in T$, we estimate:

$$d(x, (I, v)(x)) = ||v|| = ||kae_1 + lbe_2|| \geq \min\{|a|, |b|\}.$$

Denote $p_1 : \mathbb{R}^2 \to \mathbb{R}e_1$ as the projection map to the first component. Then for any $(F, ae_1/2 + v) \in Tg$ and any $x \in \mathbb{R}^2$,

$$\begin{aligned}d(x, (F, ae_1/2 + v)(x)) &= d(x, Fx + ae_1/2 + v) \\ &\geq d(p_1(x), p_1(Fx) + p_1(ae_1/2) + p_1(v)) \\ &= d(p_1(x), p_1(x) + ae_1/2 + p_1(v)) \\ &\geq d(0, ae_1/2 + p_1(v)) \\ &\geq |a|/2.\end{aligned}$$

□

3.3.3. **A small final step.** Now we have classified all free and discrete isometric action on $\mathbb{R}^2$. Recall that in Chapter 2, we postponed the proof of Lemma 2.3.14. The Lemma says that there is a uniform $D > 0$ such that $d(gx, x) \geq D$ for all $g \in \Gamma - \{e\}$ and all $x \in \mathbb{R}^2$. We need this Lemma to assure that the quotient space $\mathbb{R}^2/\Gamma$ is locally Euclidean (see proof of Theorem 2.3.15). Now we prove this Lemma.

Inspecting the last paragraph of the proof of Proposition 3.3.5 and Proposition 3.3.3 respectively, we already proved that for type IIb and IIIb. For instance, in type IIIb, we have $d(gx, x) \geq D$ for all $g \in \Gamma - \{e\}$ and all $x \in \mathbb{R}^2$, where $D = \min\{|a|/2, |b|\}$. The remaining type IIa and IIIa are easier since they are just translations. To sum up, we can choose $D > 0$ for each type as below:
- Type IIa: $D = |a|$;
- Type IIb: $D = |a|/2$;
- Type IIIa: $D = \min\{|a|, |b|\}$;
- Type IIIb: $D = \min\{|a|/2, |b|\}$.

This completes the proof of Lemma 2.3.14 that we left.

3.3.4. **Summary.** Now we have classified all 2-dimensional locally Euclidean spaces that coming from quotient spaces of free and discrete isometric actions on $\mathbb{R}^2$. We list them as below.

Theorem 3.3.6. *Let Γ be a subgroup of* $\text{Isom}(\mathbb{R}^2)$*. Suppose that Γ-action on $\mathbb{R}^2$ is free and discrete, then Γ is one of the followings:*

Type	*Generators*	*Quotient*
I	*identity*	*Plane*
IIa	(I, ae_1)	*Cylinder*
IIb	$(F, ae_1/2)$	*Twisted cylinder*
IIIa	$(I, ae_1), (I, bv)$	*Torus*
IIIb	$(F, ae_1/2), (I, be_2)$	*Klein bottle*

The generators above are written under a suitable orthonormal coordinate $\{e_1, e_2\}$.

In the next chapter, we will prove that any 2-dimensional locally Euclidean space can be realized as a quotient of a free and discrete isometric action on $\mathbb{R}^2$.

3.4. Subgroups of $\text{Isom}(\mathbb{S}^2)$

This section will be in the form of exercises, assigned through the course.

We regard $\mathbb{S}^2$ as the unit sphere in $\mathbb{R}^3$ centered at the origin. The distance on $\mathbb{S}^2$ is given as the length of shortest path on $\mathbb{S}^2$.

Here we recall some facts from linear algebra that will be helpful.

- Any eigenvalue (real or complex) of any orthogonal matrix has norm 1.
- Let $A \in O(n)$. If V is an A-invariant linear subspace of $\mathbb{R}^n$, then $V^\perp$, the orthogonal complement of V, is also A-invariant.

Exercise 3.4.1. *Show that there is a natural isomorphism between* $\text{Isom}(\mathbb{S}^2)$ *and the orthogonal group $O(3)$.*

Exercise 3.4.2. *Let $g \in \mathrm{Isom}(\mathbb{S}^2)$. Show that under a suitable orthonormal basis, the matrix representation of g has the form*

$$\begin{pmatrix} \pm 1 & 0 & 0 \\ 0 & \cos\theta & \sin\theta \\ 0 & -\sin\theta & \cos\theta \end{pmatrix}$$

Hint: First show that it must have a real eigenvalue.

Exercise 3.4.3. *Let H be a subgroup of* $\mathrm{Isom}(\mathbb{S}^2)$ *that acts freely on $\mathbb{S}^2$. Show that H is either the identity subgroup $\{I_3\}$ or $\{\pm I_3\}$.*

These exercises explore all 2-dimensional locally spherical spaces that comes from quotients of $\mathbb{S}^2$: they are either the sphere itself or $\mathbb{S}^2/\{\pm I_3\}$. As introduced in Section 2.6, the quotient $\mathbb{S}^2/\{\pm I_3\}$ is the real projective space $\mathbb{R}P^2$.

Similar to the locally Euclidean case, any 2-dimensional locally spherical spaces indeed arises from a quotient of $\mathbb{S}^2$. Together with the result in this section, this would fully classify 2-dimensional locally spherical spaces: they are $\mathbb{S}^2$ and $\mathbb{R}P^2$.

A similar classification result can be achieved for even dimensional locally spherical spaces: they are the sphere $\mathbb{S}^{2n}$ and the real projective space $\mathbb{S}^{2n}/\{\pm I_{2n+1}\}$. Odd dimensional ones have much more examples:

Example 3.4.4. We consider $\mathbb{S}^3$ as the unit sphere in $\mathbb{R}^4$. We can identity $\mathbb{R}^4$ with $\mathbb{C}^2$, by corresponding $(x, y, u, v) \in \mathbb{R}^4$ to $(x + iy, u + iv) \in \mathbb{C}^2$. We define a S^1-action on $\mathbb{C}^2$:

$$S^1 \times \mathbb{C}^2 \to \mathbb{C}^2, \quad (e^{i\theta}, (z, w)) \mapsto (e^{i\theta} z, e^{i\theta} w).$$

This induces a free isometric S^1-action on $\mathbb{S}^3$. For any $p \in \mathbb{Z}$, we take a subgroup $\mathbb{Z}_p \leq S^1$, then we have an isometric $\mathbb{Z}_p$-action on $\mathbb{S}^3$ that is free and discrete. The corresponding quotient space $\mathbb{S}^3/\mathbb{Z}_p$ is locally spherical. When $p = 2$, it is the real projective space.

Exercise 3.4.5. *Let p, q be two coprime positive integers. We define a $\mathbb{Z}_p$-action on $\mathbb{C}^2$ as follows. Let g be a generator of $\mathbb{Z}_p$,*

$$g \cdot (z, w) = (e^{i(2\pi/p)} z, e^{i(2\pi q/p)} w).$$

This induces an isometric $\mathbb{Z}_p$-action os $\mathbb{S}^3$. Show that this $\mathbb{Z}_p$-action on $\mathbb{S}^3$ is free.

The corresponding quotient spaces in the above exercise are called lens spaces, denoted as $L(p, q)$. By construction, they are also locally spherical.

CHAPTER 4

Coverings

This chapter is optional and will not be covered in lectures.

Here is the main goal of this chapter:

Theorem 4.0.1. *Let (X, d) be a 2-dimensional locally Euclidean space. Then there is a subgroup Γ of $\mathrm{Isom}(\mathbb{R}^2)$, whose action on $\mathbb{R}^2$ is free and discrete, such that (X, d) is isometric to the quotient metric space $\mathbb{R}^2/\Gamma$.*

Together with the classification of free and discrete isometric actions on $\mathbb{R}^n$ in the last chapter, Theorem 4.0.1 fully classifies all 2-dimensional locally Euclidean spaces.

In Section 4.1, we introduce the notion of covering maps and study their properties. In Section 4.2, we show that any 2-dimensional locally Euclidean space X admits a covering space $\mathbb{R}^2$. In Section 4.3, we define the covering group associated to a covering map and finish the proof of Theorem 4.0.1.

4.1. Covering maps

Recall that for a free and discrete isometric Γ-action on $\mathbb{R}^2$, we know that there is $r > 0$ such that the quotient map q provides an isometry from $B_r(x)$ to $B_r(q(x))$ for all $x \in X$. We extract this property as a definition.

Definition 4.1.1. Let (X_1, d_1) and (X_2, d_2) be two length metric spaces. We say that a map $\phi : X_1 \to X_2$ is a covering map, if
(1) ϕ is surjective, and
(2) there is $s > 0$ such that for any $a \in X_1$, the map

$$\phi|_{B_s(a)} : B_s(a) \to \phi(B_s(a))$$

is an isometry onto $B_s(\phi(a))$.

We say that (X_1, d_1) is a covering space of (X_2, d_2).

Examples 4.1.2.
- The quotient map $\mathbb{R}^2 \to \mathbb{R}^2/\Gamma$, where Γ acts on $\mathbb{R}^2$ freely and discretely by isometries.
- Wrapping a circle of circumference 2π twice over a circle of circumference π.
- Wrapping $\mathbb{R}$ over a circle.

We mention that there is a more general definition of "covering maps", which does not require any metrics. Here for our purposes, we will focus on length metric spaces, actually 2-dim locally Euclidean spaces.

Proposition 4.1.3. *Let $\phi : (X_1, d_1) \to (X_2, d_2)$ be a covering map. Let $\gamma : \mathbb{R} \to X_1$ be a geodesic in X_1, then $\phi \circ \gamma$ is a geodesic in X_2.*

Proof. Let $s > 0$ be the constant in (2) of Definition 4.1.1. Fix $t \in \mathbb{R}$. We choose $\epsilon > 0$ such that
(1) $\text{Length}(\gamma|_{[t-\epsilon,t+\epsilon]}) = d(\gamma(t-\epsilon), \gamma(t+\epsilon))$, and
(2) $d(\gamma(t-\epsilon), \gamma(t+\epsilon)) < s$.
The the curve $\gamma|_{[t-\epsilon,t+\epsilon]}$ is contained in $B_s(\gamma(t-\epsilon))$. Because ϕ is an isometry when restricted to $B_s(\gamma(t-\epsilon))$. We have

$$\text{Length}(\gamma|_{[t-\epsilon,t+\epsilon]}) = \text{Length}(\phi \circ \gamma|_{[t-\epsilon,t+\epsilon]}),$$
$$d(\gamma(t-\epsilon), \gamma(t+\epsilon)) = d(\phi \circ \gamma(t-\epsilon), \phi \circ \gamma(t+\epsilon)).$$

Thus

$$\text{Length}(\phi \circ \gamma|_{[t-\epsilon,t+\epsilon]}) = d(\phi \circ \gamma(t-\epsilon), \phi \circ \gamma(t+\epsilon)).$$

□

Lemma 4.1.4. *Let $\phi : (X_1, d_1) \to (X_2, d_2)$ be a covering map between two length metric spaces. Then*

$$d_2(\phi(a), \phi(b)) \leq d_1(a, b)$$

for all $a, b \in X_1$.

Proof. Let $\gamma : I \to X_1$ be a minimal geodesic between a and b. $\phi \circ \gamma$ provides a curve between $\phi(a)$ and $\phi(b)$. We have

$$d_2(\phi(a), \phi(b)) \leq \text{Length}(\phi \circ \gamma) = \lim_{||\pi|| \to 0} \sum_i d_2(\phi(\gamma(t_i)), \phi(\gamma(t_{i+1}))),$$

where π is a partition of I. When $||\pi||$ is sufficiently small, we have

$$d_1(\gamma(t_i), \gamma(t_{i+1})) \leq s$$

for all i, where $s > 0$ as in condition (2) of Definition 4.1.1. Applying the condition (2), we have

$$d_2(\phi(\gamma(t_i)), \phi(\gamma(t_{i+1}))) = d_1(\gamma(t_i), \gamma(t_{i+1})).$$

Thus

$$d_2(\phi(a), \phi(b)) \leq \lim_{||\pi|| \to 0} \sum_i d_2(\gamma(t_i), \gamma(t_{i+1})) = d_1(a, b).$$

□

Lemma 4.1.5. *Let $\phi : (X_1, d_1) \to (X_2, d_2)$ be a covering map between two length metric spaces. Suppose that ϕ is injective, then ϕ is an isometry.*

Proof. By assumptions, ϕ is a bijection. We need to show that ϕ preserves the distance. By Lemma 4.1.4, we have one side inequality

$$d_2(\phi(a), \phi(b)) \leq d_1(a, b).$$

To prove the other direction, we claim that the inverse $\phi^{-1} : X_2 \to X_1$ is also a covering map. Assuming this claim, then for $\phi(a) = x$ and $\phi(b) = y$, we obtain

$$d_1(a, b) = d_1(\phi^{-1}(x), \phi^{-1}(y)) \leq d_2(x, y) = d_2(\phi(a), \phi(b)).$$

This proves that ϕ is an isometry.

It remains to verify the claim. We already knew that ϕ^{-1} is surjective. We check condition (2) in Definition 4.1.1. Because ϕ is a covering map, there is $s > 0$ such that ϕ maps to $\phi(B_s(a)) = B_s(\phi(a))$ as an isometry. For a point $y \in X_2$, pick $a \in X$ with $\phi(a) = y$. It is clear that $\phi^{-1}|_{B_s(y)}$ preserves the distance. Compositing ϕ^{-1} to both sides of $\phi(B_s(a)) = B_s(\phi(a))$, we have $B_s(a) = \phi^{-1}(B_s(\phi(a)))$, that is, $B_s(\phi^{-1}(y)) = \phi^{-1}(B_s(y))$. This verifies condition (2). □

4.2. Cover a 2-dimensional locally Euclidean space by $\mathbb{R}^2$

Now we consider 2-dimensional locally Euclidean spaces, though all results can be generalized to n-dimensional ones after some mild modifications. Our goal of this section is the Proposition below.

Proposition 4.2.1. *Let X be a 2-dimensional locally Euclidean space. Then there is a covering map $\phi : \mathbb{R}^2 \to X$.*

In this section, we always assume that (X, d) is a 2-dimensional locally Euclidean space without mentioning.

Recall that a covering map sends geodesics to geodesics (Proposition 4.1.3). Now we make use of geodesics to construct covering maps and prove Proposition 4.2.1.

Lemma 4.2.2. *Let $\gamma : [a, b] \to X$ be a minimal geodesic. Then γ can be extended to a geodesic $\tilde{\gamma} : \mathbb{R} \to X$. Moreover, this extension is unique.*

Proof. First notice that for a small piece of minimal geodesic $\gamma : [b - \epsilon, b] \to B_r^2(0)$ with $\gamma(b) = 0$, we can slightly extend the domain of γ beyond b by length r, by simply extending the segment in $B_r^2(0)$. For a general locally Euclidean space, let $r > 0$ such that $B_r(x)$ is isometric to $B_r^2(0)$ for all $x \in X$. We extend the geodesic γ piece by piece, every time inside some $B_r(x)$, which is isometric $B_r^2(0)$.

Suppose that γ has two extensions γ_1 and γ_2. They agree on $[a, b]$. Without lose of generality, we assume that $\gamma_1(t_0) \neq \gamma_2(t_0)$ for some $t_0 > b$. Let

$$s = \max\{t \geq b | \gamma_1(t) \text{ and } \gamma_2(t) \text{ agree on } [b, t]\}.$$

We write $y = \gamma_1(s) = \gamma_2(s)$. Then on the ball $B_r(y)$, we have two geodesics branching at y: γ_1 and γ_2 agree on some interval $[s - \epsilon, s]$ but disagree beyond s. This cannot happen since $B_r(y)$ is isometric to $B_r^2(0)$. □

Corollary 4.2.3. *Let ϕ and ψ be two covering maps from $\mathbb{R}^2$ to X. Suppose that ϕ and ψ agree on a r-ball in $\mathbb{R}^2$, then $\phi = \psi$ everywhere.*

Proof. We assume that $\phi = \psi$ on $B_r^2(0)$, but there is $a \in \mathbb{R}^2$ such that $\phi(a) \neq \psi(a)$. Let $x = \phi(0) = \psi(0)$. Let γ be the straight line through 0 and a in $\mathbb{R}^2$. By Lemma 4.1.3, both $\phi \circ \gamma$ and $\psi \circ \gamma$ are geodesics through x in X; moreover, they agree on a small piece around x since $\phi = \psi$ on $B_r^2(0)$. This contradicts the uniqueness part of Lemma 4.2.2. □

Lemma 4.2.4 (Extension Lemma)**.** *Let $r > 0$ such that every r-ball in X is isometric to $B_r^2(0)$. Let $x \in X$ and let $\phi : B_r^2(0) \to B_r(x)$ be an isometry, then ϕ has an extension $\tilde{\phi} : \mathbb{R}^2 \to X$ with the following properties:*
(1) $\tilde{\phi}$ maps geodesics through 0 to geodesics through x;
(2) for any $a, b \in \mathbb{R}^2$ on a geodesic through 0 with $d(a, b) < r$, then

$$d(a, b) = d(\phi(a), \phi(b)).$$

Proof. We extend ϕ by extending geodesics. For any $y \notin B_r^2(0)$, we draw the unique straight ray γ emanating from 0 through y in $\mathbb{R}^2$ with $\gamma(0) = 0$. A piece of γ lies in $B_r^2(0)$. Via the isometry ϕ, this piece is mapped to a minimal geodesic starting from x in $B_r(x)$. By Lemma 4.2.2, we can extend this minimal geodesic in $B_r(x)$ to a geodesic with domain $[0, \infty)$ in X. We call this geodesic σ with $\sigma(0) = x$. For any point z on γ with $d(z, 0) = l$, we define $\tilde{\phi}(z)$ be the unique point on σ such that the piece of σ from x to $\tilde{\phi}(z)$ has length l. In particular, this defines $\tilde{\phi}(y)$. (Alternatively, you may use lines through 0 instead of rays to define $\tilde{\phi}$.)

Property (1) follows directly from our construction. It remains to prove Property (2). Let a, b on a geodesic through 0 with $d(a, b) < r$. If $a, b \in B_r^2(0)$, then desired result follows directly from the fact that ϕ is an isometry on $B_r^2(0)$. Thus without lose of generality, we can assume that a, b belong to a ray γ emanating from 0. Let $l = d(a, 0)$. For convenience, we write σ_{z_1,z_2} as the piece of σ from z_1 to z_2. By definition, $\phi(a)$ is the point on σ with $\text{Length}(\sigma_{0,\phi(a)}) = l$, and $\phi(b)$ is the point on σ with $\text{Length}(\sigma_{0,\phi(b)}) = l + d(a, b)$. Thus

$$\text{Length}(\sigma_{\phi(a),\phi(b)}) = (l + d(a, b)) - l = d(a, b) < r.$$

Because X is locally Euclidean, every small piece of σ of length $< r$ is a minimal geodesic. Hence

$$d(\phi(a), \phi(b)) = \text{Length}(\sigma_{\phi(a),\phi(b)}) = d(a, b).$$

□

We will show that the $\tilde{\phi}$ that we constructed in Lemma 4.2.4 is indeed a covering map. We need a lemma on verifying covering maps.

Lemma 4.2.5. *Let $\phi : \mathbb{R}^2 \to X$ be a map. Suppose that there is $s > 0$ such that*

$$d(\phi(a), \phi(b)) = d(a, b)$$

for all a, b with $d(a, b) < s$, then ϕ is a covering map.

Proof. Let $r > 0$ such that every r-ball in X is isometric to $B_r^2(0)$. We put $t = \min\{r, s/2\} > 0$.

We first verify condition (2) of Definition 4.1.1, that is, $\phi|_{B_t(x)}$ is an isometry onto $B_t(\phi(x))$. Let $a, b \in B_t(x)$, then

$$d(a, b) \leq d(a, x) + d(x, b) < 2t \leq s.$$

By assumptions, we know that

$$d(\phi(a), \phi(b)) = d(a, b),$$

thus $\phi(B_t(x)) \subseteq B_t(\phi(x))$ and ϕ preserves the metric on $B_t(x)$. Hence we only need to show that ϕ maps $B_t(x)$ onto $B_t(\phi(x))$. Suppose that there is a point $z \in B_t(\phi(x))$ but not in $\phi(B_t(x))$. Because ϕ is an injection when restricted to $B_s(x)$, the image $\phi(B_r(x))$ contains at least one point other than $\phi(x)$; we pick such a point $\phi(y)$. Let $l_x = d(z, \phi(x))$ and $l_y = (z, \phi(y))$. Recall that both $B_r(x)$ and $B_r(\phi(x))$ are isometric to $B_r^2(0)$. By planar geometry, there are two points (or exactly one point) w_1 and w_2 in $B_r(x)$ such that

$$d(w_j, x) = l_x, \quad d(w_j, y) = l_y, \quad j = 1, 2.$$

A similar statement also holds on $B_r(\phi(x))$, and z is one of the two points. Because ϕ preserves the distance from $B_r(x)$ to $\phi(B_r(x))$, we have

$$d(\phi(w_j), \phi(x)) = l_x, \quad d(\phi(w_j), \phi(y)) = l_y, \quad j = 1, 2.$$

Therefore, one of the $\phi(w_i)$ must be z. A contradiction.

Next we prove that ϕ is surjective. For any point $y \in X$, we need to find a point $w \in \mathbb{R}^2$ with $\phi(w) = y$. Let $\gamma : I \to X$ be a minimal geodesic from $\phi(0)$ to y. We choose a partition $\{t_0, ..., t_n\}$ of I so that $d(\gamma(t_i), \gamma(t_{i+1})) < r$ for all $i = 0, ..., n-1$. We will find points $z_i \in \mathbb{R}^2$ successively with $\phi(z_i) = \gamma(t_i)$. We start with $\phi(0) = \gamma(t_0)$. For $\gamma(t_1) \in B_r(\gamma(t_0))$, since $\phi|_{B_r^2(0)}$ provides an isometry between $B_r^2(0)$ and $B_r(\phi(0)) = B_r(\gamma(t_0))$, there is $z_1 \in B_r^2(0)$ such that $\phi(z_1) = \gamma(t_1)$. Repeating this process over balls $B_r(\gamma(t_i))$ successively (or by induction), we can find z_{i+1} with $\phi(z_{i+1}) = \gamma(t_{i+1})$. Eventually we obtain z_n with $\phi(z_n) = \gamma(t_n) = y$. □

Proof of Proposition 4.2.1. Lemma 4.2.4 provides a map $\phi : \mathbb{R}^2 \to X$. We show that ϕ is a covering map. By Lemma 4.2.5, it suffices to find $s > 0$ so that

$$d(\phi(a), \phi(b)) = d(a, b)$$

for all a, b with $d(a, b) < s$. We pick $s = r/2$, where $r > 0$ as in Lemma 4.2.4. If both a and b belong to a geodesic through 0, then Property (2) of Lemma 4.2.2 can be directly applied.

For the remaining case, we draw two rays, γ_a and γ_b, from 0 to a and b respectively. On each ray, we evenly divide the segment from 0 to a (or to b) into k many short segments so that each segment has length $< r/2$, we call the intermediate points $m_1, m_2, ..., m_k = a$ and $n_1, n_2, ..., n_k = b$. By planar geometry, $d(m_i, n_i) < r/2$ for all i. Put $m_0 = n_0 = a$. By Property (2) of Lemma 4.2.4, we have

$$d(m_i, m_{i+1}) = d(\phi(m_i), \phi(m_{i+1})), \quad d(n_i, n_{n+1}) = d(\phi(n_i), \phi(n_{i+1})).$$

Note that $\triangle\phi(m_i)\phi(n_i)\phi(n_{i+1})$ is contained in $B_r(\phi(m_i))$; since $B_r(\phi(m_i))$ is isometric to $B_r^2(0)$, we can view the triangle as one in $\mathbb{R}^2$.

Sketch of the remaining proof: Using cosine law, successively show that

$$\triangle m_i n_i n_{i+1} \cong \triangle\phi(m_i)\phi(n_i)\phi(n_{i+1}), \quad \triangle m_i m_{i+1} n_{i+1} \cong \triangle\phi(m_i)\phi(m_{i+1})\phi(n_{i+1})$$

for all i. □

Remark 4.2.6. Here we proved that the map constructed in Lemma 4.2.4 is indeed a covering map. Recall that in Lemma 4.2.4, $\phi : \mathbb{R}^2 \to X$ is extended from a local isometry $\phi : B_r^2(0) \to B_r(x)$.

A special case is when X is $\mathbb{R}^2$. For a local isometry $B_r^2(0) \to B_r^2(x) \subset \mathbb{R}^2$, through the same procedure, we obtain a covering map $\phi : \mathbb{R}^2 \to \mathbb{R}^2$.

Exercise 4.2.7. *Prove that ϕ in the above remark is an isometry.*
Hint: By Lemma 4.1.5, it suffices to show that ϕ is injective.

4.3. Covering groups

Definition 4.3.1. Let $\phi : (X_1, d_1) \to (X_2, d_2)$ be a covering map. We define the *covering group* $Cov(\phi)$ to be

$$\{g \in \mathrm{Isom}(X_1) \mid \phi(g \cdot x) = \phi(x) \text{ for all } x \in X_1\}.$$

Lemma 4.3.2. *$Cov(\phi)$ is a group.*

Proof. Clearly the identity map of X_1 belongs to $Cov(\phi)$.

Let $g \in Cov(\phi)$. Since g is an isometry, it has an inverse g^{-1}. We check that $g^{-1} \in Cov(\phi)$. This is true because for any $x \in X_1$,

$$\phi(x) = \phi(g(g^{-1}x)) = \phi(g^{-1}x).$$

Let $g_1, g_2 \in Cov(\phi)$. For any $x \in X_1$,

$$\phi((g_1 g_2) \cdot x) = \phi(g_1(g_2 x)) = \phi(g_2(x)) = \phi(x).$$

This shows $g_1 g_2 \in Cov(\phi)$. □

Lemma 4.3.3. *Let $\phi : \mathbb{R}^2 \to X$ be a covering map. Let $s > 0$ as in Definition 4.1.1. Then for any $a \in \mathbb{R}^2$ and any non-identity element $g \in Cov(\phi)$, we have $d(g \cdot a, a) \geq s$. In particular, $Cov(\phi)$-action on $\mathbb{R}^2$ is free and discrete.*

Proof. We first prove that the statement when $ga \neq a$. By Definition 4.3.1, we have $\phi(ga) = \phi(a)$. If we write $x = \phi(a)$, then both $g \cdot a$ and a are points in the pre-image $\phi^{-1}(x)$. Since $\phi|_{B_s(a)}$ is injective, we conclude $d(g \cdot a, a) \geq s$.

Now we rule out the situation $ga = a$. Suppose that $ga = a$. Since g is not the identity map, there is a point $b \in \mathbb{R}^2$ with $gb \neq b$. Let γ be the segment between a and b. Through the isometry g, $g \circ \gamma$ is a segment between a and $g \cdot b$. By planar geometry, there is a point b' on γ that is close to a such that $d(gb', b') < s$. On the other hand, we know this cannot happen from the last paragraph. □

Lemma 4.3.4. *Let $\phi : \mathbb{R}^2 \to X$ be a covering map with covering group $\Gamma = Cov(\phi)$. For any $a \in \mathbb{R}^2$ and $x = \phi(a)$, the Γ-orbit at a and the pre-image $\phi^{-1}(x)$ agrees.*

Proof. By the Definition 4.3.1, it is clear that $\Gamma \cdot a \subseteq \phi^{-1}(x)$. Only the other side inclusion is nontrivial. Let $a' \in \phi^{-1}(x)$. We need to find $g \in \Gamma$ with $g(a) = a'$. Since ϕ is a covering map, there is $s > 0$ and isometries

$$f = \phi|_{B_s(a)} : B_s(a) \to B_s(x), \quad f' = \phi|_{B_s(a')} : B_s(a') \to B_s(x).$$

Then $(f')^{-1} \circ f : B_s(a) \to B_s(a')$ is an isometry between two s-balls of $\mathbb{R}^2$. By Remark 4.2.6, we can extend this local isometry to an isometry of $\mathbb{R}^2$. We call this extension g. By construction, $g(a) = a'$. We show that $g \in \Gamma$. According to Definition 4.3.1, it suffices to show that $\phi(gy) = \phi(y)$ for all $y \in X$. Note that both $\phi \circ g$ and ϕ are covering maps from $\mathbb{R}^2$ to X. Moreover, by our construction, for any $y \in B_s(a)$,

$$\phi \circ g(y) = \phi \circ ((\phi|_{B_s(a')})^{-1} \circ \phi|_{B_s(a)})(y) = \phi|_{B_s(a)}(y) = \phi(y).$$

Hence $\phi \circ g$ and ϕ agree on $B_s(a)$. By Corollary 4.2.3, $\phi \circ g = \phi$ everywhere. This shows $g \in \Gamma$ and thus $\phi^{-1}(x) \subseteq \Gamma \cdot a$. □

Proof of Theorem 4.0.1. Let (X, d) be a 2-dimensional locally Euclidean space. By Proposition 4.2.1, there is a covering map $\phi : \mathbb{R}^2 \to X$. Let Γ be the corresponding covering group. We prove that X is isometric to $\mathbb{R}^2/\Gamma$.

We construct a map $F : X \to \mathbb{R}^2/\Gamma$ as follows. Let $x \in X$ and let $a \in \phi^{-1}(x)$. We define $F(x)$ as $\bar{a}$, where $\bar{a} \in \mathbb{R}^2/\Gamma$ is the equivalent class of a. By Lemma 4.3.4, we know that $\Gamma \cdot a = \phi^{-1}(x)$. Hence F is well-defined. One way to interpret F is using the quotient map $q : \mathbb{R}^2 \to \mathbb{R}^2/\Gamma$. By the construction above, $F(x) = q(\phi^{-1}(x))$ ($\phi^{-1}(x)$ can be either understood as the pre-image, or the inverse of a local isometry around x).

We prove that F is an isometry between X and $\mathbb{R}^2/\Gamma$.

- F is onto. For any equivalent class $\bar{a} \in \mathbb{R}^2/\Gamma$, by construction, $F(\phi(a))$ has image $q(a) = \bar{a}$.
- F is one-to-one. Suppose that we have $x, y \in X$ with $F(x) = F(y)$. This means that $\phi^{-1}(x)$ and $\phi^{-1}(y)$ lie in the same Γ-orbit. By Lemma 4.3.4, we see that $x = y$.

• F is a covering map. $(\phi|_{B_s(a)})^{-1}$ is an isometry from $B_s(x)$ to $B_s(a)$. Recall that Γ-action satisfies $d(g \cdot a, a) \geq s$ for all $a \in \mathbb{R}^2$ and all $g \in \Gamma - \{e\}$. By Theorem 2.3.17, there is $s' > 0$ such that $q|_{B_{s'}(a)} : B_{s'}(a) \to B_{s'}(\bar{a})$ is an isometry. Compositing these two isometries together from $B_{s'}(x)$ to $B_{s'}(\bar{a})$, we obtain an isometry

$$(q|_{B_{s'}(a)}) \circ (\phi|_{B_{s'}(a)})^{-1} : B_{s'}(x) \to B_{s'}(\bar{a}).$$

This map is indeed $F|_{B_{s'}(x)}$ by our construction. This shows that F is a covering map.

By Lemma 4.1.5, we conclude that F is an isometry. □

To end this chapter, we combine Theorem 4.0.1 with Theorem 3.1 in Note *Free and isometric actions on* $\mathbb{R}^2$ together, as the classification result below.

Theorem 4.3.5 (Classification of 2-dimensional locally Euclidean spaces)**.** *Let X be a 2-dimensional locally Euclidean space. Then X is isometric to a quotient metric space $\mathbb{R}^2/\Gamma$, where Γ acts on $\mathbb{R}^2$ freely and discretely by isometries. Moreover, Γ is one of the following groups with the corresponding quotient as X.*

Type	*Generators*	X
I	*identity*	*Plane*
IIa	(I, ae_1)	*Cylinder*
IIb	$(F, ae_1/2)$	*Twisted cylinder*
IIIa	$(I, ae_1), (I, bv)$	*Torus*
IIIb	$(F, ae_1/2), (I, be_2)$	*Klein bottle*

where $a, b > 0$, $\{e_1, e_2\}$ a suitable orthonormal coordinate, v a unit vector independent of e_1, $F = \mathrm{diag}\{1, -1\}$.

One can follow a similar approach in Sections 4.2 and 4.3, with some mild modifications, to prove that any 2-dimensional locally spherical space can be covered by the standard sphere $\mathbb{S}^2$. Then we can conclude the classification of these spaces as well.

Theorem 4.3.6 (Classification of 2-dimensional locally spherical spaces)**.** *Let X be a 2-dimensional locally spherical space. Then X is isometric to either $\mathbb{S}^2$ or the real projective plane $\mathbb{R}P^2 = \mathbb{S}^2/\mathbb{Z}_2$, where $\mathbb{Z}_2$ acts on $\mathbb{S}^2$ by sending a point to its antipodal point.*

CHAPTER 5

Hyperbolic Plane

5.1. Introduction

We look for a geometry that satisfies the first four axioms for the Euclidean plane geometry and a different 5th axiom (parallel postulate):
• Given a point not on a given line, there are at least two lines through the given point that does not intersect the given line.

We also wish geometry to have a large symmetry group.

Definition 5.1.1. A metric space (X, d) is homogeneous, if for any $x, y \in X$, there is $g \in \mathrm{Isom}(X)$ with $g \cdot x = y$.

A different name that commonly used is saying that $\mathrm{Isom}(X)$-action is *transitive* on X.

For example, sphere and the Euclidean plane are homogeneous.

Recall that we have seen a non-Euclidean metric on $\mathbb{R}_+ = (0, \infty)$ defined by $d(x, y) = |\ln(x/y)|$.

Proposition 5.1.2. *$(\mathbb{R}_+, d)$ is homogeneous.*

Proof. We consider the $\mathbb{R}_+$-action on $\mathbb{R}_+$ by multiplication:

$$\mu : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}_+, \quad (a, x) \to ax.$$

This action is by isometries since

$$d(ax, ay) = |\ln((ax)/ay)| = |\ln(x/y)| = d(x, y).$$

Then $(\mathbb{R}_+, d)$ is clearly homogeneous since for any $x, y \in \mathbb{R}_+$, we have

$$\mu(y/x, x) = (y/x) \cdot x = y.$$

□

Also recall that in Exercise 2.3.3, we have seen a $SL_2(\mathbb{R})$-action on $\mathbb{H}$, where $\mathbb{H} = \{z \in \mathbb{C} | \mathrm{Im}(z) > 0\}$:

$$\mu : SL_2(R) \times \mathbb{H} \to \mathbb{H}, \quad \left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}, z \right) \mapsto \frac{az + b}{cz + d}.$$

Note that for the vertical half line $\{iy | y > 0\} \subset \mathbb{H}$, the action by $\mathrm{diag}\{a, 1/a\}$ leaves this half line invariant:

$$\begin{pmatrix} a & 0 \\ 0 & 1/a \end{pmatrix} \cdot (iy) = \frac{a(iy)}{1/a} = a^2(iy).$$

We will find a distance function d on $\mathbb{H}$ so that
• $(\mathbb{H}, d)$ is a length metric space;
• the imaginary half line $\{iy | y > 0\}$ is a geodesic;
• $SL_2(\mathbb{R})$-action on $\mathbb{H}$ is isometric;

- $(\mathbb{H}, d)$ is homogeneous;
- $(\mathbb{H}, d)$ satisfies the modified 5th axiom.

5.2. Pure imaginary half line under the $SL_2(\mathbb{R})$-action

We study the $SL_2(\mathbb{R})$-action on $\mathbb{H}$ by tracking the pure imaginary half line $L = \{iy \mid y > 0\}$ under the action.

We first point out a caveat here: different matrices could correspond to the same map of $\mathbb{H}$. For example, both I_2 and $-I_2$ correspond to the identity map $z \mapsto z$. In general, A and $-I \cdot A$ also corresponds to the same map. Thus when we write the composition of maps into matrix multiplication, we may have to add an additional $-I$ (or equivalently, change the signs of all entries in a matrix) to get an equality.

We check all the elements in $SL_2(\mathbb{R})$ that leaves L invariant.

Lemma 5.2.1. *Suppose that $A \in SL_2(\mathbb{R})$ maps L to L. Then*

$$A = \begin{pmatrix} a & 0 \\ 0 & 1/a \end{pmatrix} \text{ or } \begin{pmatrix} 0 & -b \\ 1/b & 0 \end{pmatrix}.$$

Proof. By direct calculation, $A \cdot (iy)$ has real part

$$\frac{bd + acy^2}{d^2 + c^2y^2}.$$

By assumption, this real part should be zero. Note that this is impossible if $ac \neq 0$ since y could be any positive real number. If $ac = 0$, then $bd = 0$.

Case 1: $a = 0$. Because $\det(A) = 1$ and $bd = 0$, we have $d = 0$ and $c = -1/b$.

Case 2: $c = 0$. By the same reasons, we end in $c = 0$ and $d = 1/a$. □

Now we keep track of (L, d) under $SL_2(\mathbb{R})$-action. For $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, if $c = 0$ (then $ad = 1$),

$$A \cdot z = \frac{az + b}{d} = \frac{a}{d}z + \frac{b}{d} = a^2 z + \frac{b}{d}.$$

This is can be realized as a composition a multiplication (by a real number) $z \mapsto a^2 \cdot z$ and a translation (by a real number) $z \mapsto z + b/d$ and . The multiplication and the translation correspond to the matrix $\begin{pmatrix} a & 0 \\ 0 & 1/a \end{pmatrix}$ and $\begin{pmatrix} 1 & b/d \\ 0 & 1 \end{pmatrix}$, respectively. In terms of matrix multiplication, the composition says

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

Under these two elements (and their compositions), it's clear that L moves to other vertical half lines. If $c \neq 0$

$$\begin{aligned} A \cdot z = \frac{az + b}{cz + d} &= \frac{\frac{a}{c}(cz + d) + b - ad/c}{cz + d} \\ &= \frac{a}{c} + (b - ad/c) \cdot \frac{1}{az + d} \\ &= \frac{a}{c} + (bc - ad) \cdot \frac{1}{c^2 z + cd} \\ &= \frac{a}{c} + \frac{-1}{c^2 z + cd}. \end{aligned}$$

We write the above as a composition of maps:

$$z \overset{c^2\cdot}{\longmapsto} c^2 z \overset{+cd}{\longmapsto} c^2 z + cd \longmapsto \frac{-1}{c^2 z + cd} \overset{+a/c}{\longmapsto} \frac{a}{c} + \frac{-1}{c^2 z + cd}.$$

Besides the translation and multiplication (by real numbers) we have seen in the case $c = 0$ above, it has a new map involved in the composition: $z \mapsto -1/z$, which corresponds to the matrix $C = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$. (Here we cannot use $z \mapsto 1/z$ because $1/z$ would be outside $\mathbb{H}$.) In terms of matrix multiplication, we can check

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

This proves an algebraic result on $SL_2(\mathbb{R})$ by writing $z \mapsto A \cdot z$ into a composition of maps.

Proposition 5.2.2. *$SL_2(\mathbb{R})$ is generated by the matrices of the forms below:*

$$\begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} a & 0 \\ 0 & 1/a \end{pmatrix}, \quad \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix},$$

where $a \in \mathbb{R}$.

We note that C switches the first and second quadrant in $\mathbb{H}$: for $z = re^{i\theta}$,

$$C \cdot z = -\frac{1}{z} = -\frac{1}{r}e^{-i\theta} = \frac{1}{r}e^{i(\pi-\theta)}.$$

Lemma 5.2.3. *C maps the vertical half line $L_a = \{a + iy \mid y > 0\}$, where $a \neq 0$, to the upper semi-circle centered at $-1/(2a) \in \mathbb{C}$ with radius $1/|2a|$.*

Proof. For $z = a + iy$, we calculate

$$C(a+iy) = -\frac{1}{a+iy} = -\frac{a-iy}{a^2+y^2}.$$

Then we calculate its distance square to the point $-1/(2a)$:

$$\begin{aligned} d_{\mathbb{R}^2}(C(z), -1/(2a)) &= \left(-\frac{a}{a^2+y^2} + \frac{1}{2a}\right)^2 + \left(\frac{y}{a^2+y^2}\right)^2 \\ &= \frac{1}{4a^2}. \end{aligned}$$

Thus all point in $C(L_a)$ has equal distance $1/|2a|$ to the point $-1/(2a)$. Since we already we know that $C(L_a)$ is contained in $\mathbb{H}$, we see that $C(L_a)$ is in the described semi-circle. To see that $C(L_a)$ equals this semi-circle, note that

$$\lim_{y \to 0^+} C(a+iy) = (-1/a, 0), \quad \lim_{y \to +\infty} C(a+iy) = (0,0).$$

These two are exactly the two endpoints of the semi-circle on the real axis. Since $C(L_a)$ is connected, we conclude that $C(L_a)$ is the described semi-circle. ☐

Under

$$A \cdot z = \frac{a}{c} + \frac{-1}{c^2 z + cd},$$

by Lemma 5.2.3, $-1/(c^2(iy) + cd)$ form the upper semi-circle centered at $-1/(2cd)$ with radius $1/|2cd|$. It is clear that under the remaining translation $z \mapsto z + a/c$, this semi-circle moves to the upper semi-circle centered at $-1/(2cd) + a/c$ with radius $1/|2cd|$. Because a, c, d are independent parameters, these covers all the possible upper semi-circles with center on the real axis.

We denote $\mathcal{L}$ as the set of all vertical half lines $\{a+iy|y>0\}$ and all upper semi-circles orthogonal to the real axis in $\mathbb{H}$. We call any element in $\mathcal{L}$ as a hyperbolic line in $\mathbb{H}$. Summarizing the above, we have

Proposition 5.2.4. *Under $SL_2(\mathbb{R})$-action, L moves to a hyperbolic line. Moreover, any hyperbolic line is $A \cdot L$ for some $A \in SL_2(\mathbb{R})$.*

Corollary 5.2.5. *For any two hyperbolic lines in $\mathbb{H}$, there is $A \in SL_2(\mathbb{R})$ such that $A \cdot l_1 = l_2$.*

Proof. By Proposition 5.2.4, there is $B_1, B_2 \in SL_2(\mathbb{R})$ such that $B_1 \cdot L = l_1$ and $B_2 \cdot L = l_2$. Then $(B_2 B_1^{-1}) \cdot l_1 = l_2$. □

Exercise 5.2.6. *Every hyperbolic line separates $\mathbb{H}$ into two (open and connected) components; we call each component a hyperbolic half-plane. Show that for any two hyperbolic half-planes in $\mathbb{H}$, there is $A \in SL_2(\mathbb{R})$ that maps one to the other.*
Hint: Note that the matrix C switches the first and second quadrants.

5.3. Distance function on $\mathbb{H}$

We first study the distance function on the pure imaginary half line L. By Lemma 5.2.1, there are two types of matrices that leave L invariant. For $A = \begin{pmatrix} a & 0 \\ 0 & 1/a \end{pmatrix}$, we have $A \cdot (iy) = i(a^2 y)$. Suppose that $SL_2(\mathbb{R})$-action is isometric, then

$$d(iy_1, iy_2) = d(A \cdot (iy_1), A \cdot (iy_2)) = d(i(a^2 y_1), i(a^2 y_2)).$$

For $A = \begin{pmatrix} 0 & -b \\ 1/b & 0 \end{pmatrix}$, we have $A \cdot (iy) = -b^2/(iy)$. This can be broken into a composition of two maps: first take the negative inverse of iy, then multiply by b^2. In terms of matrix multiplication, we have

$$\begin{pmatrix} 0 & -b \\ 1/b & 0 \end{pmatrix} = \begin{pmatrix} b & 0 \\ 0 & 1/b \end{pmatrix} \cdot \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

We again write $C = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$. If $SL_2(\mathbb{R})$-action is isometric, then

$$d(iy_1, iy_2) = d(C(iy_1), C(iy_2)) = d(i/y_1, i/y_2).$$

Proposition 5.3.1. *Let d be a continuous distance function on L. Suppose that*
(1) the segment between two points in L is the minimal geodesic between them,
(2) the maps

$$m_a : L \to L, iy \mapsto i(ay), \quad C : L \to L, iy \mapsto -1/(iy) = i/y$$

are isometries of $(\mathbb{R}_+, d)$, where $a > 0$.
Then there is $c > 0$ such that

$$d(iy_1, iy_2) = c \cdot |\ln(y_1/y_2)|.$$

Proof. We assumption (2), we require d to satisfy

$$d(iy_1, iy_2) = d(iay_1, iay_2), \quad d(iy_1, iy_2) = d(i/y_1, i/y_2).$$

We write

$$y_1 = e^{t_1}, \quad y_2 = e^{t_2}, \quad a = e^s, \quad \rho(t_1, t_2) = d(ie^{t_1}, ie^{t_2}).$$

Then the requirements translates to

$$\rho(t_1, t_2) = \rho(s + t_1, s + t_2), \quad \rho(t_1, t_2) = \rho(-t_1, -t_2).$$

By assumption (1),

$$d(iy_1, iy_3) = d(iy_1, iy_2) + d(iy_2, iy_3),$$

where $y_1 \le y_2 \le y_3$. For $y_j = e^{t_j}$, we see

$$\rho(t_1, t_3) = \rho(t_1, t_2) + \rho(t_2, t_3),$$

where $t_1 \le t_2 \le t_3$. Let $c = d(ie, i) = \rho(1, 0) > 0$. Then for any integers $p, q > 0$,

$$q \cdot \rho(1/q, 0) = \sum_{j=0}^{q-1} \rho(j/q, (j+1)/q) = \rho(1, 0) = c;$$

$$\rho(p/q, 0) = \sum_{j=0}^{p-1} \rho(j/q, (j+1)/q) = p \cdot \rho(1/q, 0) = c \cdot p/q.$$

By the continuity of d, we conclude that for all $t > 0$, $\rho(t, 0) = c \cdot t$. For $t < 0$,

$$\rho(t, 0) = \rho(-t, 0) = c \cdot |t|.$$

For $t_1, t_2 \in \mathbb{R}$,

$$\rho(t_1, t_2) = \rho(t_1 - t_2, 0) = c \cdot |t_1 - t_2|.$$

Thus

$$d(iy_1, iy_2) = \rho(t_1, t_2) = c \cdot |t_1 - t_2| = c \cdot |\ln y_1 - \ln y_2| = c \cdot |\ln(y_1/y_2)|.$$

□

For the rest of this note, we will always use $c = 1$, thus $d(iy_1, iy_2) = |\ln(y_1/y_2)|$.

To find the distance between two arbitrary points in $\mathbb{C}$, we make use of the $SL_2(\mathbb{R})$-action and Proposition 5.2.4. Assuming that a distance function $d_{\mathbb{H}}$ on $\mathbb{H}$ satisfies

- L is a geodesic,
- $SL_2(\mathbb{R})$-action is isometric.

Then $SL_2(\mathbb{R})$-action maps L to other geodesics in $(\mathbb{H}, d_{\mathbb{H}})$. By Proposition 5.2.4, we conclude that any hyperbolic line is a geodesic.

Lemma 5.3.2. *Given any two different points $z_1, z_2 \in \mathbb{H}$, there is a unique hyperbolic line through z_1 and z_2.*

Proof. Planar geometry. □

We write the hyperbolic line through z_1 and z_2 as $l_{z_1 z_2}$. If we can find an explicit element $A \in SL_2(\mathbb{R})$ that maps $l_{z_1 z_2}$ to L, then we can use the distance function on L to define the distance between z_1 and z_2. To avoid confusions, we will write d_L as the distance function on the pure imaginary half line L, as the desired distance function on $\mathbb{H}$ as $d_{\mathbb{H}}$.

If $\mathrm{Re}(z_1) = \mathrm{Re}(z_2) = a$, then $l_{z_1z_2}$ is the vertical half line through $\{a+iy|y>0\}$. Clearly the translation $A : z \mapsto z - a$ maps $l_{z_1z_2}$ to L. Since A is an isometry and $d_{\mathbb{H}} = d_L$ on L by our assumption, we have

$$d_{\mathbb{H}}(z_1, z_2) = d_L(z_1 - a, z_2 - a) = \left|\ln \frac{z_1 - a}{z_2 - a}\right|.$$

Since $z_j - a \in L$, an equivalent way to write this is

$$d_{\mathbb{H}}(z_1, z_2) = \left|\ln \frac{|z_1 - a|}{|z_2 - a|}\right|,$$

where $|z - a|$ means the Euclidean distance between z and a.

If $\mathrm{Re}(z_1) \neq \mathrm{Re}(z_2)$, then $l_{z_1z_2}$ is a semicircle orthogonal to the real axis. Let $a < b$ be the two endpoints of this semicircle. Then claim that

$$A : z \mapsto \frac{z - b}{z - a}$$

maps $l_{z_1z_2}$ to L. This map does not look like in $SL_2(\mathbb{R})$ at first sight, but notice that it can be rewritten as

$$\frac{z - b}{z - a} = \frac{cz - cb}{cz - ca},$$

where $c = 1/\sqrt{b - a}$. Then the matrix

$$\begin{pmatrix} c & -cb \\ c & -ca \end{pmatrix}$$

has determinant 1. Now we verify the claim. Note that when $z = b$, $A \cdot z = 0$. Hence $A \cdot l_{z_1,z_2}$ should have 0 as a endpoint. Also, as $z \to a$, $A \cdot z$ has norm $\to \infty$. Hence $A \cdot l_{z_1,z_2}$ cannot be bounded. We also know that $A \cdot l_{z_1,z_2}$ is a hyperbolic line. Combining all these facts together, we conclude that $A \cdot l_{z_1,z_2}$ is L. This gives the distance formula:

$$d_{\mathbb{H}}(z_1, z_2) = d_L(A \cdot z_1, A \cdot z_2) = d\left(\frac{z_1 - b}{z_1 - a}, \frac{z_2 - b}{z_2 - a}\right) = \left|\ln\left(\frac{z_1 - b}{z_1 - a}\Big/\frac{z_2 - b}{z_2 - a}\right)\right|.$$

There is a convenient way to rewrite this. Note that $(z_j - b)/(z_j - a)$ is on the pure imaginary half line, thus

$$\frac{z_1 - b}{z_1 - a}\Big/\frac{z_2 - b}{z_2 - a} = \left|\frac{z_1 - b}{z_1 - a}\right|\Big/\left|\frac{z_2 - b}{z_2 - a}\right| = \frac{|z_1 - b|}{|z_1 - a|}\Big/\frac{|z_2 - b|}{|z_2 - a|}.$$

With this observation, we can write

$$d_{\mathbb{H}}(z_1, z_2) = \left|\ln\left(\frac{|z_1 - b|}{|z_1 - a|}\Big/\frac{|z_2 - b|}{|z_2 - a|}\right)\right|.$$

Again, we understand $|z - a|$ as the Euclidean distance between z and a.

Exercise 5.3.3. *Let l be the upper semi-circle in $\mathbb{H}$ with radius 1 and center 0.*
(1) Find a matrix $P \in SL_2(\mathbb{R})$ that maps l to L.
(2) Find all matrices in $SL_2(\mathbb{R})$ that preserves l.
Hint: Note that A preserves L if and only if $P^{-1}AP$ preserves l.

Lemma 5.3.4. *$d_{\mathbb{H}}$ defined above is well-defined (independent of $A \in SL_2(\mathbb{R})$ that maps l_{z_1,z_2} to L).*

PROOF. To show that $d_{\mathbb{H}}$ on $\mathbb{H}$ is well-defined, suppose that $A, B \in SL_2(\mathbb{R})$ both maps $l_{z_1.z_2}$ to L, we need to prove that

$$d_L(A \cdot z_1, A \cdot z_2) = d_L(B \cdot z_1, B \cdot z_2).$$

In fact, note that the composition $BA^{-1} \in SL_2(\mathbb{R})$ maps L to itself. Also recall that d_L is chosen so that it is invariant under any matrix in $SL_2(\mathbb{R})$ that preserves L. Hence BA^{-1} is an isometry of (L, d_L). It follows that

$$d_L(A \cdot z_1, A \cdot z_2) = d_L(BA^{-1} \cdot A \cdot z_1, BA^{-1} \cdot A \cdot z_2) = d_L(B \cdot z_1, B \cdot z_2).$$

□

Lemma 5.3.5. *$SL_2(\mathbb{R})$-action on $\mathbb{H}$ preserves $d_{\mathbb{H}}$, that is,*

$$d_{\mathbb{H}}(z_1, z_2) = d_{\mathbb{H}}(A \cdot z_1, A \cdot z_2)$$

for all $z_1, z_2 \in \mathbb{H}$ and all $A \in SL_2(\mathbb{R})$.

PROOF. Let l be the hyperbolic line through z_1, z_2 and let $B \in SL_2(\mathbb{R})$ mapping l to L. Similarly, we let l' be the hyperbolic line through Az_1, Az_2 and let $B' \in SL_2(\mathbb{R})$ mapping l' to L. By the definition of $d_{\mathbb{H}}$, it suffices to show that

$$d_L(Bz_1, Bz_2) = d_L(B'Az_1, B'Az_2).$$

Note that

$$B \cdot l = L, \quad A \cdot l = l', \quad B' \cdot l' = L.$$

Thus $B'AB^{-1} \in SL_2(\mathbb{R})$ maps L to itself. Therefore,

$$d_L(Bz_1, Bz_2) = d(B'AB^{-1}Bz_1, B'AB^{-1}Bz_2) = d_L(B'Az_1, B'Az_2).$$

□

Proposition 5.3.6. *$d_{\mathbb{H}}$ satisfies the triangle inequality.*

We will prove Proposition 5.3.6 in the next section. Assuming that it holds at this moment, we derive the theorem below.

Theorem 5.3.7. *$(\mathbb{H}, d_{\mathbb{H}})$ is a homogeneous length metric space.*

PROOF. It follows from the definition of $d_{\mathbb{H}}$ and Proposition 5.3.6 that $d_{\mathbb{H}}$ is a distance function. Lemma 5.3.5 says that $SL_2(\mathbb{R})$-action on $(\mathbb{H}, d_{\mathbb{H}})$ is isometric.

We show that $(\mathbb{H}, d_{\mathbb{H}})$ is a length metric space. Let $z_1, z_2 \in \mathbb{H}$ and let l be the hyperbolic line through them. Let γ be a piece of this hyperbolic line with endpoints z_1 and z_2. By Corollary, we can find $A \in SL_2(\mathbb{R})$ with $A \cdot l = L$. Then $A \cdot \gamma$ is a segment in L between Az_1 and Az_2. Recall that (L, d_L) is a length metric space as we showed in Section 2.1. Together with the isometric $SL_2(\mathbb{R})$-action, we conclude that

$$dd_{\mathbb{H}}(z_1, z_2) = d_L(Az_1, Az_2) = \text{length}(A \cdot \gamma) = \text{length}(\gamma).$$

Thus γ realizes the distance between z_1 and z_2.

Now we prove the homogeneous part. It suffices to show that for any $z \in \mathbb{H}$, we can find a matrix in $SL_2(\mathbb{R})$ that maps z to i. Let l be a hyperbolic line through z. By Corollary 5.2.5, there is $A \in SL_2(\mathbb{R})$ with $A \cdot l = L$. Then $A \cdot z \in L$. Further composing a multiplication by a positive number, we send z to i. □

Exercise 5.3.8. *Let (X, d) be a metric space. We say that (X, d) is two-point homogeneous, if for any two pairs of points A_1, B_1 and A_2, B_2 with $d(A_1, B_1) = d(A_2, B_2)$, there is an isometry g of (X, d) such that $g(A_1) = A_2$ and $g(B_1) = B_2$.*
Show that $(\mathbb{H}, d_{\mathbb{H}})$ is two-point homogeneous.

To close this section, we note that since geodesics in $(\mathbb{H}, d_{\mathbb{H}})$ are hyperbolic lines, this geometry satisfies the modified parallel postulate: given a hyperbolic line l and a point z outside l, there are at least two (actually infinitely many) hyperbolic lines that go through z and does not intersect l.

5.4. Projection

Let $d = d_{\mathbb{H}}$ be defined as in the last section. To prove the triangle inequality (Proposition 5.3.6), we will introduce the notion of projection.

Let $z_1, z_2, z_3 \in \mathbb{H}$. We want to show that

$$d(z_1, z_2) \leq d(z_1, z_3) + d(z_3, z_2).$$

If these three points belong to a common hyperbolic line l, then we can use some $A \in SL_2(\mathbb{R})$ that maps l to L and then apply the triangle inequality on (L, d_L). Regarding the general case, after applying some $A \in SL_2(\mathbb{R})$, we can assume without lose of generality that $z_1, z_2 \in L$ and $z_3 \notin L$. We will find a *projection* of z_3 to L, which we write as z_3', such that

$$d(z_3', iy) \leq d(z_3, iy)$$

for any point $iy \in L$. If such a projection z_3' exists, then triangle inequality follows:

$$d(z_1, z_2) \leq d(z_1, z_3') + d(z_3', z_2) \leq d(z_1, z_3) + d(z_3, z_2).$$

To keep the notations concise, we will write w as a point outside L.

Definition 5.4.1. Let $w \in \mathbb{H} - L$. We define its projection to L as follows: we draw the upper semi-circle (hyperbolic line) l centered at 0 and through w. This semi-circle intersects L perpendicularly at a point w'. We call w' the projection of w to L.

Proposition 5.4.2. *Let $w \in \mathbb{H} - L$ and let $w' \in L$ be the projection of w to L. Then $d(w', z) < d(w, z)$ for all $z \in L$.*

Proof. We prove the case that $|z| \geq |w'|$ here. The other case $|z| \leq |w'|$ is similar with some mild modifications.

We draw a hyperbolic line through w and z. We denote the endpoints of this upper semi-circle as a and b on $\mathbb{R}$ (we denote a as the end point in the same quadrant as w). By the definition of d, we have

$$d(w, z) = \left| \ln \left(\frac{|w - b|}{|w - a|} \Big/ \frac{|z - b|}{|z - a|} \right) \right| = \left| \ln \frac{\tan \theta_w}{\tan \theta_z} \right| = \ln \frac{\tan \theta_z}{\tan \theta_w},$$

where $\theta_w = \angle wba$ and $\theta_z = \angle zba$ (the last equality above follows from $\theta_z > \theta_w$). Also,

$$d(w', z) = \left| \ln \frac{|w'|}{|z|} \right| = \ln \frac{|z|}{|w'|}.$$

It suffices to show that

$$\frac{\tan \theta_z}{\tan \theta_w} > \frac{|z|}{|w'|}.$$

Let $\theta' = \angle w'ba$. Note that

$$\tan\theta_z = \frac{|z|}{|b|}, \quad \tan\theta_w < \tan\theta' = \frac{|w'|}{|b|}.$$

This implies the desired inequality. □

We can interpret the inequality in Proposition 5.4.2 in a different way.

Corollary 5.4.3. *Let l be the upper semi-circle with center* 0 *and radius R. Let w be a point on l and z be a point on L. Then*

$$d(w', z) = \min_{w\in l} d(w, z).$$

Moreover, w' is the unique point in l that obtains the minimum.

Proof. Note that as w changes in l, by the construction in Definition 5.4.1, w' does not change. Thus Proposition 5.4.2 implies that

$$d(w', z) \le d(w, z)$$

for all $w \in l$, where the equality holds if and only if $w = w'$. The result follows. □

One may ask whether w' is the closest point to w on L. Or more generally, one may ask given a hyperbolic line l and a point z outside l, which point on l has the shortest distance to z.

To answer these questions, we will make use of $SL_2(\mathbb{R})$-action and Corollary 5.4.3. We need to first address the concern whether being perpendicular is preserved under $SL_2(\mathbb{R})$-action.

Let l_1 and l_2 be two hyperbolic lines in $\mathbb{H}$ intersecting at a point z_0. We define the angle between these two hyperbolic lines as their angle in $\mathbb{R}^2$. More precisely, we can parameterize these two curves as γ_1 and γ_2 with $\gamma_1(0) = \gamma_2(0) = z_0$, then we just use the angle between the tangent vectors $\gamma_1'(0)$ and $\gamma_2'(0)$ (we denote this as $\angle(\gamma_1'(0), \gamma_2'(0))$ for convenience). More generally, we can also use the same idea to define the angles between two C^1 curves in $\mathbb{H}$.

We show that $SL_2(\mathbb{R})$-action is angle-preserving.

Theorem 5.4.4. *For any $A \in SL_2(\mathbb{R})$, the map $F : z \mapsto Az$ preserves the angles. In other words, for any two vectors v_1 and v_2 based at any point $z \in \mathbb{H}$, we have*

$$\angle(v_1, v_2) = \angle(DF_z(v_1), DF_z(v_2)),$$

where DF_z is the differential of F at z, which is a 2×2 matrix.

Note that $\angle(v_1, v_2)$ is based at z, while $\angle(DF_z(v_1), DF_z(v_2))$ is based at $F(z)$.

We usually call angle-preserving maps as *conformal maps.* For example, the following maps of $\mathbb{R}^2$ are conformal:

- A translation $(x, y) \mapsto (x + a, y + b)$.
- An isometry of $(\mathbb{R}^2, d)$, where d is the standard Euclidean metric.
- A scaling (multiplication) $(x, y) \mapsto (ax, ay)$, where $a > 0$.
- A composition of the above maps.

Proof of Theorem 5.4.4. Since the composition of two conformal maps is also conformal, by Proposition 5.2.2, it suffices to show that the three types of generators of $SL_2(\mathbb{R})$ preserve the angle. They are

$$z \mapsto z + a, \quad z \mapsto a^2 z, \quad z \mapsto -1/z,$$

where $a \in \mathbb{R}$. We check the last one, since the first two are quite clear.

For F with $F(z) = -1/z$, to calculate its differential as a 2×2 matrix, write F in Cartesian coordinate:

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

thus we treat this map as

$$F : \mathbb{R}^2 \to \mathbb{R}^2, \quad (x,y) \mapsto \left(\frac{-x}{x^2+y^2}, \frac{y}{x^2+y^2} \right).$$

Then its differential at (x, y) is

$$DF_{(x,y)} = \begin{pmatrix} \frac{x^2-y^2}{(x^2+y^2)^2} & \frac{2xy}{(x^2+y^2)^2} \\ \frac{-2xy}{(x^2+y^2)^2} & \frac{x^2-y^2}{(x^2+y^2)^2} \end{pmatrix}.$$

Note that

$$\left(\frac{x^2-y^2}{(x^2+y^2)^2} \right)^2 + \left(\frac{2xy}{(x^2+y^2)^2} \right)^2 = \frac{1}{(x^2+y^2)^2}.$$

Thus we can rewrite the differential as

$$DF_{(x,y)} = \begin{pmatrix} \frac{1}{x^2+y^2} & 0 \\ 0 & \frac{1}{x^2+y^2} \end{pmatrix} \cdot \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}.$$

for some θ. The first matrix in the product corresponds to a multiplication (scaling), while the second one corresponds to a rotation. Both matrices preserve the angles, thus its product $DF_{(x,y)}$ also preserves the angles. □

Theorem 5.4.5. *Given a hyperbolic line l and a point $z \in \mathbb{H}$ outside l, then there is a unique point z' on l such that*

$$d(z, z') = \min_{w \in l} d(z, w).$$

This z' can be constructed as follows: through z, we draw a hyperbolic line that is perpendicular to l, then the intersection point is z'.

We call this point z' the *projection* of z to l. In terms of Corollary 5.4.3, we think of w' as the projection of z to l. Note that the construction here also matches with Definition 5.4.1, though for different purposes.

Proof. Let l_0 be the unit upper semi-circle centered at 0. By Corollary 5.2.5, there is $A \in SL_2(\mathbb{R})$ that maps l to l_0.

If $Az \in L$, then we directly apply Corollary 5.4.3 to conclude that the distance between Az and l_0 is obtained only at point p, the intersection of L and l_0:

$$d(Az, p) = \min_{w \in l} d(Az, Aw).$$

Let $z' = A^{-1}p$, then

$$d(z, z') = d(Az, p) = \min_{w \in l} d(Az, Aw) = \min_{w \in l} d(z, w).$$

Moreover, since A preserves angles and maps hyperbolic lines to hyperbolic lines, z' is the intersection of l and a hyperbolic line that is through z and is perpendicular to l.

If $Az \notin L$, we will compose one more isometry B so that $BAz \in L$ and B preserves l_0, then the argument in the last paragraph applies. To find this $B \in SL_2(\mathbb{R})$, recall that according to Exercise 5.3.3, we can consider the matrices of the form

$$P^{-1}\begin{pmatrix} a & 0 \\ 0 & 1/a \end{pmatrix} P, \quad \text{where } P = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}.$$

Since any point on L can be expressed as $\operatorname{diag}\{a, 1/a\} \cdot i$ for some $a > 0$, we can parameterize l_0 as

$$\gamma(t) = P^{-1}\begin{pmatrix} t & 0 \\ 0 & 1/t \end{pmatrix} P \cdot i = P^{-1} \cdot it^2, \quad t \in (0, \infty).$$

It suffices to show that the path

$$\sigma(t) = P^{-1}\begin{pmatrix} t & 0 \\ 0 & 1/t \end{pmatrix} P \cdot (Az)$$

intersects L. Note that

$$d(\gamma(t), \sigma(t)) = d(i, Az),$$

which has fixed value for all $t > 0$. As $t \to 0$, $\gamma(t) \to 1$; thus $\sigma(t)$ belongs to the first quadrant for all t sufficiently small (actually, $\sigma(t) \to 1$ as $t \to 0$). Similarly, since $\gamma(t) \to -1$ as $t \to \infty$, $\sigma(t)$ lies in the second quadrant for all t large. By intermediate value theorem, there is t_0 such that $\sigma(t_0) \in L$. □

5.5. The isometry group of $(\mathbb{H}, d)$

Other than isometries from the $SL_2(\mathbb{R})$-action, $(\mathbb{H}, d)$ actually has other isometries.

Lemma 5.5.1. *$h(z) = -\bar{z}$ is an isometry of $(\mathbb{H}, d)$.*

Proof. It is clearly that h maps $\mathbb{H}$ to $\mathbb{H}$ as a bijection. To show that h preserves the metric, note that h maps $z = x + iy$ to $-\bar{z} = -x + iy$. Thus h is the reflection about the pure imaginary axis and preserve the Euclidean distance. Thus the hyperbolic distance is also preserved under h due to the definition of d. □

Note that h above is also conformal, since in Cartesian coordinate, h is a reflection $(x, y) \mapsto (-x, y)$.

Lemma 5.5.2. *For $a, b, c, d \in \mathbb{R}$ with $ad - bc = -1$, the map*

$$g(z) = \frac{a\bar{z} + b}{c\bar{z} + d}$$

is an isometry of $(\mathbb{H}, d)$.

Proof. If $c = 0$, then $ad = -1$ and

$$g(z) = \frac{a\bar{z} + b}{d} = \frac{a}{d}\bar{z} + \frac{b}{d} = -a^2\bar{z} + \frac{b}{d}.$$

g is composition of maps

$$z \overset{h}{\mapsto} -\bar{z} \overset{a^2\cdot}{\mapsto} -a^2\bar{z} \overset{+b/d}{\mapsto} g(z)$$

with each map is an isometry.

If $c \neq 0$, we recycle the calculation in a previous section to obtain

$$g(z) = \frac{a}{c} + (bc - ad) \cdot \frac{1}{c^2\bar{z} + cd} = \frac{a}{c} + \frac{1}{c^2\bar{z} + cd}$$

We can decompose g as

$$z \overset{h}{\mapsto} -\bar{z} \overset{c^2\cdot}{\mapsto} -c^2\bar{z} \overset{-cd}{\mapsto} -c^2\bar{z} - cd \overset{-1/}{\mapsto} \frac{1}{c^2\bar{z} + cd} \overset{+a/c}{\mapsto} g(z).$$

Each map involved is an isometry, thus the result follows. □

We define a set G to be all the isometries of $(\mathbb{H}, d)$ that we have seen so far. They are in the following forms:

$$z \mapsto \frac{az + b}{cz + d} \text{ with } ad - bc = 1; \quad \text{or} \quad z \mapsto \frac{a\bar{z} + b}{c\bar{z} + d} \text{ with } ad - bc = -1.$$

They are called isometries of the first kind and the second kind, respectively.

Isometries of the first kind is usually called *Mobuis transformations*. One minor caveat here: the set of Mobuis transformations is not one-to-one correspond to $SL_2(\mathbb{R})$, because I_2 and $-I_2$ indeed correspond to the same transformation. If you have learned about group theory, then you know something called *quotient group*. The group consisting of all Mobuis transformations is the so-called *projective special linear group*

$$PSL_2(\mathbb{R}) = SL_2(\mathbb{R})/\{\pm I_2\}.$$

Isometries of the second kind do not form a group. In fact, the identity element is not included, and it is not closed under composition: let $h(z) = -\bar{z}$, then $h^2 = \text{id}$.

Exercise 5.5.3. *Show that G is a group.*

One can prove this exercise by showing the composition of maps of G still has the required form. In fact, if g_1, g_2 are of first kind and h_1, h_2 are of second kind, then one can easily check

- $g_1 g_2$ and $h_1 h_2$ are of first kind;
- $g_1 h_1$ and $h_1 g_1$ are of second kind.

Next we show that G is indeed the entire isometry group.

Theorem 5.5.4. *$G = \text{Isom}(\mathbb{H}, d)$.*

We prove two lemmas first.

Lemma 5.5.5. *Let g be an isometry of $(\mathbb{H}, d)$ that preserves L and fixes two distinct points in L, then g fixes every point in L.*

PROOF. Let z_1 and z_2 be two fixed points of g on L. Note that for any point w on L, w is uniquely determined by its distance to z_1 and z_2:

$$r_1 = d(z_1, w), \quad r_2 = d(z_2, w).$$

$g(w)$ satisfies

$$d(g(w), z_i) = d(g(w), g(z_i)) = d(w, z_i) = r_i,$$

where $i = 1, 2$. Hence $g(w) = w$ for all $w \in L$. □

Lemma 5.5.6. *Let g be an isometry of $(\mathbb{H}, d)$ such that $g(iy) = iy$ for all $y > 0$. Then either g is the identity map or g is $z \mapsto -\bar{z}$.*

PROOF. Let $z \in \mathbb{H} - L$ and let z' be its projection to L. By Theorem 5.4.5, z' satisfies

$$d(z, z') = \min_{w \in L} d(z, w).$$

Applying the isometry g, we have

$$d(g(z), g(z')) = \min_{w \in L} d(g(z), g(w)).$$

Since g fixes every point on L, this turns to

$$d(g(z), z') = \min_{w \in L} d(g(z), w).$$

It follows that z' is the projection of $g(z)$ to L. By the construction of projection as described in Theorem 5.4.5, we conclude that z, z', and $g(z)$ all belong to the hyperbolic line that is perpendicular to L at z'. Together with

$$d(z, z') = d(g(z), z'),$$

we see that $g(z)$ is either z or $-\bar{z}$.

Let Q_1 and Q_2 be the first (open) quadrant and the second quadrant in $\mathbb{H}$. Since g is a continuous bijection of $\mathbb{H}$ that fixes the pure imaginary half line L, it must preserve each Q_i $(i = 1, 2)$ or switch Q_1 and Q_2.

Case 1: g preserves Q_1 and Q_2. Then $g(z) = z$ for all $z \in \mathbb{H}$.

Case 2: g switches Q_1 and Q_2. Then $g(z) = -\bar{z}$ for all $z \in \mathbb{H}$. □

We are ready to prove Theorem 5.5.4.

Proof of Theorem 5.5.4. Let g be an isometry of $(\mathbb{H}, d)$. We need to show that $g \in G$. We will write g as a composition of isometries of the first or second kind.

Let z_1 and z_2 be two distinct points in L. Then $g(z_1)$ and $g(z_2)$ is a pair of points in $\mathbb{H}$ such that

$$d(z_1, z_2) = d(g(z_1), g(z_2)).$$

By your work in Exercise 5.3.8, there is $A \in SL_2(\mathbb{R})$ such that $Az_1 = g(z_1)$ and $Az_2 = g(z_2)$; moreover, $A \cdot L = l$, where $l = g(L)$ is the hyperbolic line through $g(z_1)$ and $g(z_2)$. Thus $A^{-1}g$ is an isometry that preserves L and fixes $z_1, z_2 \in L$. By Lemmas 5.5.5 and 5.5.6, $A^{-1}g$ is either the identity map or is $h \in G$ defined by $h(z) = -\bar{z}$. Thus either $g = A$ or $g = Ah$. This completes the proof. □

To have a better understanding of the isometries, we take a look at their fixed points.

Lemma 5.5.7. *Let g be a non-identity isometry of the first kind, then either g has exactly one fixed point or no fixed points.*

Proof. We solve the following equation in $\mathbb{H}$

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

that is,

$$az + b = cz^2 + dz,$$

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

When $c = 0$, we get $(d - a)z = b$. If $d = a$ and $b = 0$, then g is the identity map. If $d = a$ not $b \neq 0$, then equation $(d - a)z = b$ has no solutions and thus g has no fixed points. If $d \neq a$, then $(d - a)z = b$ has no solutions in $\mathbb{H}$.

When $c \neq 0$, this is a quadratic equation with real coefficients. It has exactly two roots in $\mathbb{C}$ counting multiplicity. If two roots are real, then it has no solutions in $\mathbb{H}$; also note that this happens if and only if

$$0 \leq (d-a)^2 + 4bc = d^2 + a^2 - 2ad + 4bc = (d+a)^2 - 4(ad-bc) = (d+a)^2 - 4.$$

If we have complex roots, then exactly one of them is in $\mathbb{H}$ since these two roots are conjugating to each other. □

Definition 5.5.8. We say two elements h_1, h_2 in a group G are *conjugate* to each other, if there is $g \in G$ such that $g^{-1}h_1 g = h_2$.

For example, let $h_1 : z \mapsto z+1$ and $h_a z \mapsto z + a$, where $a \in \mathbb{R} - \{0\}$. With $g : z \mapsto a^{-1}z$, we have

$$g^{-1}h_1 g(z) = g^{-1}h_1(a^{-1}z) = g^{-1}(a^{-1}z + 1) = z + a = h_a(z).$$

Thus h_a is conjugate to h_1.

Remark 5.5.9. The kind of g does not change under a conjugation.

Proposition 5.5.10. *Let g be an isometry of the first kind that has no fixed points in $\mathbb{H}$, then one of the followings holds for g:*
(1) g is conjugate to $z \mapsto az$, where $a \in (0,1) \cup (1, \infty)$;
(2) g is conjugate to $z \mapsto z+1$.

Proof. According to the proof of Lemma 5.5.7, there are four cases that g has no fixed points in $\mathbb{H}$.
Case 1. $c = 0$, $d = a$, and $b \neq 0$. Then g is the map $z \mapsto z \pm b$, which is conjugate to $z \mapsto z+1$.
Case 2. $c = 0$ and $d \neq a$. Then it is direct to check that g preserves the vertical line with real part $b/(d-a)$ (this value comes from the solution to $g(x) = x$):

$$g\left(\frac{b}{d-a} + iy\right) = \frac{a}{d} \cdot \left(\frac{b}{d-a} + iy\right) + \frac{b}{d},$$

which has real part

$$\frac{a}{d} \cdot \frac{b}{d-a} + \frac{b}{d} = \frac{b}{d-a}.$$

We will continue on this case later.
Case 3. $c \neq 0$ and $a + d = \pm 2$. When $a + d = \pm 2$, let λ_1, λ_2 be two eigenvalues of the matrix A with entries a, b, c, d. Then

$$\lambda_1 + \lambda_2 = \mathrm{tr} A = \pm 2, \quad \lambda_1 \lambda_2 = \det A = 1.$$

We see that $\lambda_1 = \lambda_2 = \pm 1$. By linear algebra, we can find $P \in GL_2(\mathbb{R})$ such that

$$P^{-1}AP = \begin{pmatrix} \pm 1 & b' \\ 0 & d' \end{pmatrix}.$$

Adjusting P by scaling all entries by a constant, we can assume that $P \in SL_2(\mathbb{R})$. Since this matrix has determinant 1, d' has to be ± 1. Also note that $b' \neq 0$; otherwise $A = \pm I_2$ but we assumed $c \neq 0$. Since the matrix $\begin{pmatrix} \pm 1 & b' \\ 0 & \pm 1 \end{pmatrix}$ correspond to the map $z \mapsto z \pm b'$, we end in (2) of the statement.
Case 4. $c \neq 0$ and $a + d = \in (-2, 2)$. We show that g preserves a hyperbolic line in this case. To find the desired hyperbolic line, we regard g as a map with domain $D = \mathbb{H} \cup \mathbb{R}$. The quadratic equation in Lemma 5.5.7 has two distinct real roots

x_1, x_2; they are fixed under g. Recall that g sends hyperbolic lines to hyperbolic lines, thus g must preserve the upper semi-circle with endpoints x_1, x_2.

In both cases 2 and 4, we have g preserves some hyperbolic line l. Let $P \in SL_2(\mathbb{R})$ that maps l to L. We note that $A \in SL_2(\mathbb{R})$ maps L to L if and only if AP maps l to L, if and only if $P^{-1}AP$ maps l to l. Thus g is conjugate to an isometry described in Lemma 5.2.1, that is, g can be expressed as

$$P^{-1}\begin{pmatrix} a & 0 \\ 0 & 1/a \end{pmatrix} P \quad \text{or} \quad P^{-1}\begin{pmatrix} 0 & -b \\ 1/b & 0 \end{pmatrix} P.$$

It is direct to check that the second type above indeed has a fixed point as $P^{-1} \cdot |b|i$, thus this case cannot happen. Therefore, we end in (1) of the statement. □

For isometries of the first kind that have exactly one fixed point, one can prove the result below:

Exercise 5.5.11. *(1) Find all the matrices $A \in SL_2(\mathbb{R})$ with $A \cdot i = i$.*
(2) Show that an isometry of first kind has at least one fixed point if and only if it is conjugate to the isometry represented by the matrix

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

We apply the same approach to understand isometries of the second kind.

Proposition 5.5.12. *Any isometry of the second kind has the set of fixed points as a hyperbolic line or the empty set.*

Proof. We solve the equation

$$g(z) = \frac{a\bar{z} + b}{c\bar{z} + d} = z.$$

$$a\bar{z} + b = c\bar{z}z + dz = c|z|^2 + dz.$$

Write $z = x + iy$. Then the imaginary parts of the above equality gives $-ay = dy$. If $a \neq -d$, then there are no fixed points in $\mathbb{H}$.

If $a = -d$ and $c = 0$, we have

$$a\bar{z} + b = dz, \quad a(\bar{z} + z) = -b, \quad 2\mathrm{Re}(z) = -b/a.$$

This represents a vertical half-line with real part equal to $-b/(2a)$.

If $a = -d$ and $c \neq 0$, we write $\alpha = a/c$ and $\beta = b/c$, then

$$\alpha\bar{z} + \beta = |z|^2 - \alpha z, \quad |z|^2 - \alpha z - \alpha\bar{z} + \alpha^2 = \beta + \alpha^2,$$

$$(z - \alpha)(\bar{z} - \alpha) = \beta + \alpha^2, \quad |z - \alpha|^2 = \beta + \alpha^2.$$

Note that

$$\beta + \alpha^2 = (a/c)^2 + (b/c) = (a^2 + bc)/c^2 = (-ad + bc)/c^2 = 1/c^2 > 0.$$

Thus $|z - \alpha|^2 = \beta + \alpha^2$ represents a semi-circle centered at α with radius $1/c$, which is the set of fixed points. □

Corollary 5.5.13. *Let g be an isometry of the second kind that has the set of fixed points as a hyperbolic line. Then g is conjugate to $z \mapsto -\bar{z}$.*

Proof. Suppose that g fixes all the points on some hyperbolic line l. Let $P \in SL_2(\mathbb{R})$ that maps l tp L. Then PgP^{-1} fixes every point on L: in fact, for any $P(z) \in L$, where $z \in l$, we have

$$PgP^{-1}(P(z)) = Pg(z) = P(z).$$

Thus g is conjugate to an isometry that fixes L. The results now follows from Lemma 5.5.6. □

Exercise 5.5.14. *Let g be an isometry of the second kind that has no fixed points.*
(1) Show that g must preserve a hyperbolic line.
(2) Show that g is conjugate to $z \mapsto -a\bar{z}$, where $a \in (0,1) \cup (1,\infty)$.

We take a closer look at the isometry that fixes every point on a hyperbolic line. We continue to use the notations in the proof of Proposition 5.5.12.

If $c = 0$, we write $\gamma = -b/(2a)$, then

$$g(z) = \frac{a\bar{z} + b}{d} = -\bar{z} - \frac{b}{a} = 2\gamma - \bar{z}.$$

$\bar{z} + g(z) = 2\gamma$ means that g is the reflection about the vertical line with real part γ (you can draw a picture to see this).

If $c \neq 0$, then

$$g(z) = \frac{\alpha\bar{z} + \beta}{\bar{z} - \alpha} = \alpha + \frac{\alpha^2 + \beta}{\bar{z} - \alpha} = \alpha + \frac{1/c^2}{\bar{z} - \alpha}.$$

Thus

$$(g(z) - \alpha)(\bar{z} - \alpha) = 1/c^2.$$

Because the right hand side is a real number, we can rewrite the equality as

$$|g(z) - \alpha| \cdot |z - \alpha| = 1/c^2.$$

Moreover, α, z and $g(z)$ belong to the same line: we show that $(g(z) - \alpha)/(z - \alpha)$ is a real number:

$$\frac{g(z) - \alpha}{z - \alpha} = \frac{1/c^2}{(\bar{z} - \alpha)(z - \alpha)} = \frac{1}{c^2|z - \alpha|^2} \in \mathbb{R}.$$

This map g is called the *inversion* about the circle with radius $1/c$ and center α.

We summarize these above results as the following corollary.

Corollary 5.5.15. *Let g be an isometry of $(\mathbb{H}, d)$. Then g is conjugate to one of the followings:*

$$z \mapsto z + 1, \quad z \mapsto az, \quad z \mapsto -a\bar{z}, \quad z \mapsto \frac{\cos\theta \cdot z + \sin\theta}{-\sin\theta \cdot z + \cos\theta},$$

where $a > 0$, $\theta \in (0, 2\pi)$.