This is supporting material for the course. However the material of record for the course, for which you are responsible, is what was actually covered in class. Make sure that you take good notes.
CHAPTER 1

Tool kit

This chapter is a quick list of notation and basic mathematical definitions used in the course. Do not read this yet. We will return to these sections when they are needed.

1.1. The language of set theory

For us, a set $X$ is a collection of objects called its elements. In theory, one needs to be more careful in the definition of sets in order to avoid logical inconsistencies. However, these deep and subtle issues do not arise at the level of the mathematics described in this book. We will consequently be content with the above intuitive definition.

When the object $x$ is an element of the set $X$, we say that $x$ belongs to $X$ and we write $x \in X$.

In practice, a set can be described by listing all of its elements between curly brackets, or by describing a property that characterizes the elements of the set. For instance, the set of all even integers that are strictly between $-3$ and $8$ is $\{-2, 0, 2, 4, 6\}$ or $\{x; \text{there exists an integer } n \text{ such that } x = 2n \text{ and } -3 < x < 8\}$.

A subset of a set $X$ is a set $Y$ such that every element of $Y$ is also an element of $X$. We then write $Y \subset X$.

A particularly useful set is the empty set $\emptyset = \{\}$, which contains no element.

A bar across a symbol indicates that the corresponding property does not hold. For instance, $x \notin X$ means that $x$ does not belong to the set $X$.

Here is a list of classical sets of numbers, with the notation used in this book:

- $\mathbb{N} = \{1, 2, 3, \ldots \}$ is the set of all positive integers; in particular, $0$ is not an element of $\mathbb{N}$, a convention which is not universal.
- $\mathbb{Z} = \{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots \}$ is the set of all integers.
- $\mathbb{Q}$ is the set of all rational numbers, namely of all numbers that can be written as a quotient $\frac{p}{q}$ where $p, q \in \mathbb{Z}$ are integers with $q \neq 0$.
- $\mathbb{R}$ is the set of all real numbers.
- $\mathbb{C}$ is the set of all complex numbers. (See Section 1.2 later in this tool kit.)

Given two sets $X$ and $Y$, their intersection $X \cap Y$ consists of all elements that are in both $X$ and $Y$. Their union $X \cup Y$ consists of those elements that are in $X$ or in $Y$ (or in both). The complement $X - Y$ consists of those elements of $X$ which do not belong to $Y$. For instance, if $X = \{1, 2, 3, 4\}$ and $Y = \{3, 4, 5\}$, then

\[
X \cap Y = \{3, 4\},
\]

\[
X \cup Y = \{1, 2, 3, 4, 5\}
\]

and $X - Y = \{1, 2\}$.

Two sets are disjoint when their intersection is empty.
We can also consider the product of $X$ and $Y$, which is the set $X \times Y$ consisting of all ordered pairs $(x, y)$ where $x \in X$ and $y \in Y$. More generally, the product $X_1 \times X_2 \times \cdots \times X_n$ of $n$ sets $X_1$, $X_2$, \ldots, $X_n$ consists of all ordered $n$–tuples $(x_1, x_2, \ldots, x_n)$ where each coordinate $x_i$ is an element of $X_i$.

In particular,
\[
\mathbb{R}^2 = \mathbb{R} \times \mathbb{R} = \{(x, y); x \in \mathbb{R}, y \in \mathbb{R}\}
\]
is naturally identified to the plane through cartesian coordinates. The same holds for the 3–dimensional space
\[
\mathbb{R}^3 = \mathbb{R} \times \mathbb{R} \times \mathbb{R} = \{(x, y, z); x \in \mathbb{R}, y \in \mathbb{R}, z \in \mathbb{R}\}.
\]

A map or function $\varphi: X \to Y$ is a rule $\varphi$ which to each $x \in X$ associates an element $\varphi(x) \in Y$. We also express this by saying that the map $\varphi$ is defined by $x \mapsto \varphi(x)$. Note the slightly different arrow shape.

When $X = Y$, there is a special map, called the identity map $\operatorname{Id}_X: X \to X$ which to each $x \in X$ associates itself, namely such that $\operatorname{Id}_X(x) = x$ for every $x \in X$.

The composition of two maps $\varphi: X \to Y$ and $\psi: Y \to Z$ is the map $\psi \circ \varphi: X \to Z$ defined by the property that $\psi \circ \varphi(x) = \psi(\varphi(x))$ for every $x \in X$. In particular, $\varphi = \operatorname{Id}_Y \circ \varphi = \varphi \circ \operatorname{Id}_X$ for any function $\varphi: X \to Y$.

The map $\varphi$ is injective or one-to-one if $\varphi(x) \neq \varphi(x')$ for every $x, x' \in X$ with $x \neq x'$. It is surjective or onto if every $y \in Y$ is the image $y = \varphi(x)$ of some $x \in X$. The map $\varphi$ is bijective if it is both injective and surjective, namely if every $y \in Y$ is the image $y = \varphi(x)$ of a unique $x \in X$. In this case, there is a well-defined inverse map $\varphi^{-1}: Y \to X$, for which $\varphi^{-1}(y)$ is the unique $x \in X$ such that $y = \varphi(x)$. In particular, $\varphi \circ \varphi^{-1} = \operatorname{Id}_Y$ and $\varphi^{-1} \circ \varphi = \operatorname{Id}_X$. When $\varphi: X \to Y$ is bijective, we also say that it defines a one-to-one correspondence between elements of $X$ and elements of $Y$.

The image of a subset $A \subset X$ under the map $\varphi: X \to Y$ is the subset
\[
\varphi(A) = \{y \in Y; y = \varphi(x) \text{ for some } x \in X\}
\]
of $Y$. The preimage of $B \subset Y$ under $\varphi: X \to Y$ is the subset
\[
\varphi^{-1}(B) = \{x \in X; \varphi(x) \in B\}
\]
of $X$. Note that the preimage $\varphi^{-1}$ is defined even when $\varphi$ is not bijective, in which case the inverse map $\varphi^{-1}$ may not be defined and the preimage of a point may be empty or consist of many points.

The map $\varphi: X \to X$ preserves or respects a subset $A \subset X$ if $\varphi(A)$ is contained in $A$. A fixed point for $\varphi$ is an element $x \in X$ such that $\varphi(x) = x$; equivalently, we then say that $\varphi$ fixes $x$.

If we have a map $\varphi: X \to Y$ and a subset $A \subset X$, the restriction of $\varphi$ to $A$ is the function $\varphi|_A: A \to X$ defined by restricting attention to elements of $A$, namely defined by the property that $\varphi|_A(a) = \varphi(a)$ for every $a \in A$.

When the map $\varphi: \mathbb{N} \to X$ is defined on the set $\mathbb{N}$ of all positive integers, it is called a sequence. In this case, it is customary to write $\varphi(n) = P_n$ (with the integer $n$ as a subscript) and to denote the sequence by a list $P_1, P_2, \ldots, P_n, \ldots$, or by $(P_n)_{n \in \mathbb{N}}$ for short.

### 1.2. Complex numbers

In the plane $\mathbb{R}^2$, we can consider the $x$–axis $\mathbb{R} \times \{0\}$ as a copy of the real line $\mathbb{R}$, by identifying the point $(x, 0) \in \mathbb{R} \times \{0\}$ to the number $x \in \mathbb{R}$. If we set $i = (0, 1)$, then every
point of the plane can be written as a linear combination \((x, y) = x + iy\). When using this notation, we will consider \(x + iy\) as a generalized number, called a \textit{complex number}. It is most likely that the reader already has some familiarity with complex numbers, and we will just review a few of their properties.

Complex numbers can be added in the obvious manner
\[
(x + iy) + (x' + iy') = (x + x') + i(y + y'),
\]
and multiplied according to the rule that \(i^2 = -1\), namely
\[
(x + iy)(x' + iy') = (xx' - yy') + i(x'y + xy').
\]
These addition and multiplication behave according to the standard rules of algebra. For instance, given three complex numbers \(z = x + iy\), \(z' = x' + iy'\) and \(z'' = x'' + iy''\), we have
\[
z(z' + z'') = zz' + zz''
\]
and
\[
z(z'z'') = (zz')z''
\]
For a complex number \(z = x + iy\), the \(x\)-coordinate is called the \textit{real part} \(\text{Re}(z) = x\) of \(z\), and the \(y\)-coordinate is its \textit{imaginary part} \(\text{Im}(z) = y\). The \textit{complex conjugate} of \(z\) is the complex number
\[
\bar{z} = x - iy
\]
and the \textit{modulus}, or \textit{absolute value} of \(z\) is
\[
|z| = \sqrt{x^2 + y^2} = \sqrt{zz}.
\]
In particular,
\[
\frac{1}{x + iy} = \frac{\bar{z}}{zz} = \frac{\bar{z}}{|z|^2} = \frac{x}{x^2 + y^2} - i\frac{y}{x^2 + y^2}.
\]
Also,
\[
\bar{zz'} = (xx' - yy') - i(xy' + yx') = (x - iy)(x' - iy') = \bar{z}\bar{z'}
\]
and
\[
|zz'| = \sqrt{zz'\bar{z}\bar{z'}} = \sqrt{zz}\sqrt{zz'} = |z||z'|
\]
for every \(z = x + iy\) and \(z' = x' + iy' \in \mathbb{C}\).

In the book, we make extensive use of Euler’s \textit{exponential notation}, where
\[
\cos \theta + i\sin \theta = e^{i\theta}
\]
for every \(\theta \in \mathbb{R}\). In particular, any complex number \(z = x + iy\) can be written as \(z = r e^{i\theta}\), where \([r, \theta]\) are polar coordinates describing the same point \(z\) as the cartesian coordinates \((x, y)\) in the plane \(\mathbb{R}^2\).

There are many ways to justify this exponential notation. For instance, one can remember the Taylor expansions
\[
\sin \theta = \sum_{k=0}^{\infty} (-1)^k \frac{\theta^{2k+1}}{(2k + 1)!} = \theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} - \frac{\theta^7}{7!} + \ldots
\]
\[
\cos \theta = \sum_{k=0}^{\infty} (-1)^k \frac{\theta^{2k}}{(2k)!} = 1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - \frac{\theta^6}{6!} + \ldots
\]
\[
e^{\theta} = \sum_{n=0}^{\infty} \frac{\theta^n}{n!} = 1 + \theta + \frac{\theta^2}{2!} + \frac{\theta^3}{3!} + \frac{\theta^4}{4!} + \frac{\theta^5}{5!} + \frac{\theta^6}{6!} + \frac{\theta^7}{7!} + \ldots
\]
valid for every $\theta \in \mathbb{R}$. If, symbolically, we replace $\theta$ by $i\theta$ in the last equation and remember that $i^2 = -1$,

$$e^{i\theta} = 1 + (i\theta) + \frac{(i\theta)^2}{2!} + \frac{(i\theta)^3}{3!} + \frac{(i\theta)^4}{4!} + \frac{(i\theta)^5}{5!} + \frac{(i\theta)^6}{6!} + \frac{(i\theta)^7}{7!} + \ldots$$

$$= 1 + i\theta - \frac{\theta^2}{2!} - i\frac{\theta^3}{3!} + \frac{\theta^4}{4!} + i\frac{\theta^5}{5!} - \frac{\theta^6}{6!} - i\frac{\theta^7}{7!} + \ldots$$

$$= (1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - \frac{\theta^6}{6!} + \ldots) + i(\theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} - \frac{\theta^7}{7!} + \ldots)$$

$$= \cos \theta + i \sin \theta.$$

There is actually a way to justify this symbolic manipulation by proving the absolute convergence of this infinite series of complex numbers.

In the same vein, using the addition formulas for trigonometric functions,

$$e^{i\theta}e^{i\theta'} = (\cos \theta + i \sin \theta)(\cos \theta' + i \sin \theta')$$

$$= (\cos \theta \cos \theta' - \sin \theta \sin \theta') + i(\cos \theta \sin \theta' + \sin \theta \cos \theta')$$

$$= \cos(\theta + \theta') + i \sin(\theta + \theta')$$

$$= e^{i(\theta + \theta')},$$

which is again consistent with the exponential notation.

Note the special case

$$e^{ix} = -1,$$

known as Euler’s Formula, which combines two of the most famous mathematical constants (three if one includes the number 1 among famous constants).

1.3. Maximum, minimum, supremum and infimum

If a set $A = \{x_1, x_2, \ldots, x_n\}$ consists of finitely many real numbers, there is always one on these numbers which is larger than all the other ones, and another one which is smaller than the other ones. These are the maximum $\max A$ and the minimum $\min A$ of $A$, respectively.

However, the same does not hold for infinite subsets of $\mathbb{R}$. For instance, the set $A = \{2^n; n \in \mathbb{Z}\}$ does not have a maximum, because if contains elements that are arbitrarily large. It has no minimum either because there is no $a \in A$ such that $a \leq 2^n$ for every $n \in \mathbb{Z}$.

We can fix this problem by doing two things. First, we introduce a point $\pm \infty$ at each end of the number line $\mathbb{R} = (-\infty, +\infty)$, so as to get a new set $[-\infty, +\infty] = \mathbb{R} \cup \{-\infty, +\infty\}$. Then, we will say that an element $M \in [-\infty, +\infty]$ is a supremum for the subset $A \subset \mathbb{R}$ if:

1. $a \leq M$ for every $a \in A$;
2. $M$ is the smallest number with this property, in the sense that there is no $M' < M$ such that $a \leq M'$ for every $a \in A$.

The second condition is equivalent to the property that one can find elements of $A$ that are arbitrarily close to $M$.

Similarly, an infimum for $A \subset \mathbb{R}$ is an element $m \in [-\infty, +\infty]$ such that:

1. $a \geq m$ for every $a \in A$;
2. $m$ is the largest number with this property, in the sense that there is no $m' > m$ such that $a \geq m'$ for every $a \in A$. 
It is a deep result of real analysis that any subset $A \subset \mathbb{R}$ admits a \textit{unique supremum} $M = \sup A$, and a \textit{unique infimum} $m = \inf A$. The proof of this statement requires a deep understanding of the nature of real number. To a large extent, real numbers were precisely introduced for this property to hold true, and some people even use it as an axiom in the construction of real numbers. We refer to any undergraduate textbook on real analysis for a discussion of this statement.

For instance,

$$
\sup \{2^n; n \in \mathbb{Z}\} = +\infty \\
\text{and } \inf \{2^n; n \in \mathbb{Z}\} = 0.
$$

It may happen that $\sup A$ is an element of $A$, in which case we say that the supremum is also a \textit{maximum} and we write $\sup A = \max A$; otherwise the maximum of $A$ does not exist. Similarly, the \textit{minimum} $\min A$ of $A$ is equal to $\inf A$ if this infimum belongs to $A$, and does not exist otherwise. In particular, the maximum and the minimum are elements of $A$ when they exist. The supremum and infimum always exist, but are not necessarily in $A$.

For instance, $\min \{2^n; n \in \mathbb{N}\} = 2$, but $\min \{2^n; n \in \mathbb{Z}\}$ does not exist since $\inf \{2^n; n \in \mathbb{Z}\} = 0 \notin \{2^n; n \in \mathbb{Z}\}$.

You should beware of the behavior of suprema and infima under arithmetic operations. For instance, if we are given two sequences $(x_n)_{n \in \mathbb{N}}$ and $(y_n)_{n \in \mathbb{N}}$ of real numbers, it is relatively easy to check that

$$
\sup \{x_n + y_n; n \in \mathbb{N}\} \leq \sup \{x_n; n \in \mathbb{N}\} + \sup \{y_n; n \in \mathbb{N}\}
$$

and

$$
\inf \{x_n + y_n; n \in \mathbb{N}\} \geq \inf \{x_n; n \in \mathbb{N}\} + \inf \{y_n; n \in \mathbb{N}\}.
$$

However, these inequalities will be strict in most cases. Similarly,

$$
\sup \{-x_n; n \in \mathbb{N}\} = -\inf \{x_n; n \in \mathbb{N}\}
$$

and

$$
\inf \{-x_n; n \in \mathbb{N}\} = -\sup \{x_n; n \in \mathbb{N}\}.
$$

Finally, you may enjoy considering the case of the empty set $\emptyset$, and justify the fact that $\sup \emptyset = -\infty$ and $\inf \emptyset = +\infty$.

### 1.4. Limits and continuity. Limits involving infinity

In Section 2.3, we define limits and continuity in metric spaces by analogy with the corresponding notions that one encounters in calculus. It may be useful to review these calculus definitions.

Let $f : \mathcal{D} \to \mathbb{R}$ be a function with domain $\mathcal{D} \subset \mathbb{R}$. The function $f$ is \textit{continuous} at $x_0 \in \mathcal{D}$ if $f(x)$ is arbitrary close to $f(x_0)$ when $x \in \mathcal{D}$ is sufficiently close $x_0$. This intuitive statement is made rigorous by quantifying the adverbs “arbitrarily” and “sufficiently” with appropriate numbers $\varepsilon$ and $\delta$. In this precise definition of continuity, the function $f$ is continuous at $x_0$ if, for every $\varepsilon > 0$, there exists a number $\delta > 0$ such that $|f(x) - f(x_0)| < \varepsilon$ for every $x \in \mathcal{D}$ with $|x - x_0| < \delta$. This property is more relevant when $\varepsilon$ and $\delta$ are both small, and this is the situation that one should keep in mind to better understand the meaning of the definition.

We can reinforce the analogy with the metric space definition given in Section 2.3 by using the notation $d(x, y) = |x - y|$, namely by considering the usual metric $d$ of the real line. The above definition can then be rephrased by saying that $f$ is continuous at $x_0 \in \mathcal{D}$
if, for every $\varepsilon > 0$, there exists a $\delta > 0$ such that $d(f(x), f(x_0)) < \varepsilon$ for every $x \in D$ with $d(x, x_0) < \delta$.

Also, a sequence of real numbers $x_1, x_2, \ldots, x_n, \ldots$ converges to $x_\infty \in \mathbb{R}$ if $x_n$ is arbitrarily close to $x_\infty$ when the index $n$ is sufficiently large. More precisely, the sequence $(x_n)_{n \in \mathbb{N}}$ converges to $x_\infty$ if, for every $\varepsilon > 0$, there exists an $n_0$ such that $|x_n - x_\infty| < \varepsilon$ for every $n \geq n_0$. If, again, one replaces the statement $|x_n - x_\infty| < \varepsilon$ by $d(x_n, x_\infty) < \varepsilon$, we recognize here the definition of limits in metric spaces that is given in Section 2.3.

In calculus, one also encounters infinite limits and limits as one goes to $\pm \infty$. Recall that $f(x)$ has a limit $L \in \mathbb{R}$ as $x$ tends to $+\infty$ if, for every $\varepsilon > 0$, there exists a number $\eta > 0$ such that $|f(x) - L| < \varepsilon$ for every $x$ with $x > \eta$. Similarly, $f(x)$ converges to $L$ as $x$ tends to $-\infty$ if, for every $\varepsilon > 0$, there exists a number $\eta > 0$ such that $|f(x) - L| < \varepsilon$ for every $x$ with $x < -\eta$. In both cases, the more relevant situation is that where $\varepsilon$ is small and $\eta$ is large.

In the book, we combine $+\infty$ and $-\infty$ into a single infinity $\infty$. Then, by definition, $f(x)$ converges to $L$ as $x$ tends to $\infty$ if, for every $\varepsilon > 0$, there exists a number $\eta > 0$ such that $|f(x) - L| < \varepsilon$ for every $x$ with $|x| > \eta$.

Beware that the symbols $\infty$ and $+\infty$ represent different mathematical objects in these statements. In particular, $\lim_{x \to \infty} f(x) = L$ exactly when the properties that $\lim_{x \to +\infty} f(x) = L$ and $\lim_{x \to -\infty} f(x) = L$ both hold.

Similarly, $f(x)$ converges to $\infty$ as $x$ tends to $x_0$ if, for every number $\eta > 0$, there exists a $\delta > 0$ such that $|f(x)| > \eta$ for every $x$ with $0 < |x - x_0| < \delta$. In particular, $\lim_{x \to x_0} f(x) = \infty$ if either $\lim_{x \to x_0} f(x) = +\infty$ or $\lim_{x \to x_0} f(x) = -\infty$. However, the converse is not necessarily true, as illustrated by the fact that $\lim_{x \to 0} \frac{1}{x} = \infty$ but that neither $\lim_{x \to 0} \frac{1}{x} = +\infty$ nor $\lim_{x \to 0} \frac{1}{x} = -\infty$ hold.
CHAPTER 2

The euclidean plane

We are all very familiar with the geometry of the euclidean plane $\mathbb{R}^2$. We will encounter a new type of 2-dimensional geometry in the next chapter, that of the hyperbolic plane $\mathbb{H}^2$. In this chapter, we first list a series of well-known properties of the euclidean plane which, in the next chapter, will enable us to develop the properties of the hyperbolic plane in very close analogy.

Before proceeding forward, you are advised to briefly consult the ‘tool kit’ in the appendix for a very succinct review of the basic definitions and notation concerning set theory, infima and suprema of sets of real numbers, and complex numbers.

2.1. Euclidean length and distance

The euclidean plane is the set
$$\mathbb{R}^2 = \{(x, y); x, y \in \mathbb{R}\}$$
consisting of all ordered pairs $(x, y)$ of real numbers $x$ and $y$.

![Figure 2.1. The euclidean plane](image)

It $\gamma$ is a curve in $\mathbb{R}^2$, parametrized by the differentiable vector-valued function
$$t \mapsto (x(t), y(t)) \quad a \leq t \leq b,$$
its euclidean length $\ell_{\text{euc}}(\gamma)$ is the arc length given by

$$\ell_{\text{euc}}(\gamma) = \int_a^b \sqrt{x'(t)^2 + y'(t)^2} \, dt$$

This length is independent of the parametrization, by a well-known consequence of the chain rule.

It will be convenient to consider piecewise differentiable curves $\gamma$, made up of finitely many differentiable curves $\gamma_1, \gamma_2, \ldots, \gamma_n$ such that the initial point of each $\gamma_{i+1}$ is equal to
the end point of $\gamma_i$. In other words, such a curve $\gamma$ is differentiable everywhere except at finitely many points, corresponding to the end points of the $\gamma_i$, where it is allowed to have a "corner" (but no discontinuity). In this case, the length $\ell_{\text{euc}}(\gamma)$ of the piecewise differentiable curve $\gamma$ is defined as the sum of the lengths $\ell_{\text{euc}}(\gamma_i)$ of its differentiable pieces $\gamma_i$. This is equivalent to allowing the integrand in (2.1) to be undefined at finitely many values of $t$ where, however, it has finite left-hand and right-hand limits.

The euclidean distance $d_{\text{euc}}(P, Q)$ between two points $P$ and $Q$ is the infimum of the lengths of all piecewise differentiable curves $\gamma$ going from $P$ to $Q$, namely

$$d_{\text{euc}}(P, Q) = \inf \{ \ell_{\text{euc}}(\gamma); \gamma \text{ goes from } P \text{ to } Q \}$$

See §1.3 for basic facts about the infimum of a set of real numbers. By definition of the infimum, the above definition means that every piecewise differentiable curve $\gamma$ going from $P$ to $Q$ must have length greater than or equal to $d_{\text{euc}}(P, Q)$, and that there are curves whose length is arbitrarily close to $d_{\text{euc}}(P, Q)$.

### 2.2. Shortest curves

It is well-known and easily proved (see Exercise 2.2) that the straight line provides the shortest route between two points.

**Proposition 2.1.** The distance $d_{\text{euc}}(P, Q)$ is equal to the euclidean length $\ell_{\text{euc}}([P, Q])$ of the line segment $[P, Q]$ going from $P$ to $Q$. In other words, $[P, Q]$ is the shortest curve going from $P$ to $Q$. $\square$

In particular, computing the length of a line segment by using the formula (2.1) for arc length (see Exercise 2.1), we obtain:

**Corollary 2.2.** The euclidean distance from $P_0 = (x_0, y_0)$ to $P_1 = (x_1, y_1)$ is equal to

$$d_{\text{euc}}(P_0, P_1) = \sqrt{(x_1 - x_0)^2 + (y_1 - y_0)^2} \quad \square$$

### 2.3. Metric spaces

The euclidean plane $\mathbb{R}^2$, with its distance function $d_{\text{euc}}$, is a fundamental example of a metric space. A metric space is a pair $(X, d)$ consisting of a set $X$ together with a function $d: X \times X \to \mathbb{R}$ such that

1. $d(P, Q) \geq 0$ and $d(P, P) = 0$ for every $P, Q \in X$;
2. $d(P, Q) = 0$ if and only if $P = Q$;
3. $d(Q, P) = d(P, Q)$ for every $P, Q \in X$;
4. $d(P, R) \leq d(P, Q) + d(Q, R)$ for every $P, Q, R \in X$.

The fourth condition is the triangle inequality. The function $d$ is called the distance function, the metric function, or just the metric of the metric space $X$.

A function $d$ that satisfies only Conditions (1), (3) and (4) is called a semi-distance function or a semi-metric.

Elementary and classical properties of euclidean geometry show that $(\mathbb{R}^2, d_{\text{euc}})$ is a metric space. In particular, this explains the terminology for the triangle inequality. In fact, $(\mathbb{R}^2, d_{\text{euc}})$ and its higher dimensional analogs are typical examples of metric spaces. See Exercise 2.3 for a proof that $d_{\text{euc}}$ is a distance function which, instead of prior knowledge about euclidean geometry, uses only the definition of the euclidean distance by Equation (2.2).
The main point of the definition of metric spaces is that the notions about limits and continuity that one encounters in calculus (see Section 1.4 in the tool kit appendix) immediately extend to the wider context of a metric space \((X, d)\).

For instance, a sequence of points \(P_1, P_2, \ldots, P_n, \ldots\) in \(X\) converges to the point \(P_\infty\) if, for every \(\varepsilon > 0\), there exists an integer \(n_0\) such that \(d(P_n, P_\infty) < \varepsilon\) for every \(n \geq n_0\). This is equivalent to the property that the sequence \((d(P_n, P_\infty))\) converges to 0, as a sequence of real numbers. The point \(P_\infty\) is the limit of the sequence \((P_n)_{n \in \mathbb{N}}\).

Similarly, a function \(\varphi: X \rightarrow X'\) from a metric space \((X, d)\) to a metric space \((X', d')\) is continuous at \(P_0 \in X\) if, for every point every number \(\varepsilon > 0\), there exists a \(\delta > 0\) such that \(d'(\varphi(P), \varphi(P_0)) < \varepsilon\) for every \(P \in X\) with \(d(P, P_0) < \delta\). The function is continuous if it is continuous at every \(P_0 \in X\).

We will make extensive use of the notion of ball in a metric space \((X, d)\). The (open) ball with center \(P_0 \in X\) and radius \(r > 0\) in \((X, d)\) is the subset
\[
B_d(P_0, r) = \{P \in X; d(P, P_0) < r\}.
\]

The terminology is motivated by the case where \(X\) is the 3-dimensional euclidean space \(\mathbb{R}^3\), and where \(d\) is the euclidean metric \(d_{\text{euc}}\) defined by the property that
\[
d_{\text{euc}}(P_0, P_1) = \sqrt{(x_1 - x_0)^2 + (y_1 - y_0)^2 + (z_1 - z_0)^2}
\]
when \(P_0 = (x_0, y_0, z_0)\) and \(P_1 = (x_1, y_1, z_1)\). In this case, \(B_{d_{\text{euc}}}(P_0, r)\) is of course a geometric ball of radius \(r\) centered at \(P_0\), without its boundary.

When \((X, d)\) is the euclidean plane \((\mathbb{R}^2, d_{\text{euc}})\), a ball \(B_{d_{\text{euc}}}(P_0, r)\) is an open disk of radius \(r\) centered at \(P_0\). When \((X, d)\) is the real line \((\mathbb{R}, d)\) with its usual metric \(d(x, y) = |x-y|\), the ball \(B_{d_{\text{euc}}}(x_0, r)\) is just the open interval \((x_0-r, x_0+r)\).

Incidentally, this may be a good spot to remind the reader of a few definitions which are often confused. In the euclidean plane \(\mathbb{R}^2\), the open disk \(B_{d_{\text{euc}}}(P_0, r) = \{P \in \mathbb{R}^2; d_{\text{euc}}(P, P_0) < r\}\) is not the same thing as the circle \(\{P \in \mathbb{R}^2; d_{\text{euc}}(P, P_0) = r\}\) with center \(P_0\) and radius \(r\) that bounds it. Similarly, in dimension 3, the open ball \(B_{d_{\text{euc}}}(P_0, r) = \{P \in \mathbb{R}^3; d_{\text{euc}}(P, P_0) < r\}\) should not be confused with the sphere \(\{P \in \mathbb{R}^3; d_{\text{euc}}(P, P_0) = r\}\) with the same center and radius.

### 2.4. Isometries

The euclidean plane has many symmetries. In a metric space, these are called isometries. An isometry between two metric spaces \((X, d)\) and \((X', d')\) is a bijection \(\varphi: X \rightarrow X'\) which respects distances, namely such that
\[
d'(\varphi(P), \varphi(Q)) = d(P, Q)
\]
for every \(P, Q \in X\).

Recall that the statement that \(\varphi\) is a bijection means that \(\varphi\) is one-to-one (or injective) and onto (or surjective), so that it has a well-defined inverse \(\varphi^{-1}: X' \rightarrow X\). It immediately follows from definitions that the inverse \(\varphi^{-1}\) of an isometry \(\varphi\) is also an isometry.

It is also immediate that an isometry is continuous.

When there exists an isometry \(\varphi\) between two metric spaces \((X, d)\) and \((X', d')\), then these two spaces have exactly the same properties. Indeed, \(\varphi\) can be used to translate any property of \((X, d)\) to the same property for \((X', d')\).

We are here interested in the case where \((X, d) = (X', d') = (\mathbb{R}^2, d_{\text{euc}})\). Isometries of \((\mathbb{R}^2, d_{\text{euc}})\) include:
• **translations** along a vector \((x_0, y_0)\), defined by 
  \[\varphi(x, y) = (x + x_0, y + y_0);\]

• **rotations** of angle \(\theta\) around the origin, 
  \[\varphi(x, y) = (x \cos \theta - y \sin \theta, x \sin \theta + y \cos \theta);\]

• **reflections** across a line passing through the origin and making an angle of \(\theta\) with the \(x\)-axis, 
  \[\varphi(x, y) = (x \cos 2\theta + y \sin 2\theta, x \sin 2\theta - y \cos 2\theta);\]

• more generally, any composition of the above isometries, namely any map \(\varphi\) of the form
  \[(2.4) \quad \varphi(x, y) = (x \cos \theta - y \sin \theta + x_0, x \sin \theta + y \cos \theta + y_0)\]
  or
  \[(2.5) \quad \varphi(x, y) = (x \cos 2\theta + y \sin 2\theta + x_0, x \sin 2\theta - y \cos 2\theta + y_0).\]

For the last item, recall that the composition of two maps \(\varphi: X \to Y\) and \(\psi: Y \to Z\) is the map \(\psi \circ \varphi: X \to Z\) defined by \(\psi \circ \varphi(P) = \psi(\varphi(P))\) for every \(P \in X\).

The above isometries are better expressed in terms of complex numbers, identifying the point \((x, y)\) with the complex number \(z = x + iy \in \mathbb{C}\). See Section 1.2 in the tool kit appendix for a brief summary of the main properties of complex numbers.

Using Euler’s exponential notation (see Section 1.2)
  \[e^{i\theta} = \cos \theta + i \sin \theta,\]

the isometries listed in Equations (2.4) and (2.5) can then be written as
  \[\varphi(z) = e^{i\theta}z + z_0\]

and
  \[\varphi(z) = e^{2i\theta}\bar{z} + z_0\]

where \(z_0 = x_0 + iy_0\) and where \(\bar{z} = x - iy\) is the complex conjugate of \(z = x + iy\).

**Proposition 2.3.** If \(\varphi\) is an isometry of \((\mathbb{R}^2, d_{\text{eucl}}) = (\mathbb{C}, d_{\text{eucl}})\), then there exists a point \(z_0 \in \mathbb{C}\) and an angle \(\theta \in \mathbb{R}\) such that
  \[\varphi(z) = e^{i\theta}z + z_0 \quad \text{or} \quad \varphi(z) = e^{2i\theta}\bar{z} + z_0\]

for every \(z \in \mathbb{C}\).

**Proof.** See Exercise 3.3 (and compare Theorem 3.11) in Chapter 3 for a proof of this well-known result in euclidean geometry.

A fundamental consequence of the abundance of isometries of the euclidean plane \((\mathbb{R}^2, d_{\text{eucl}})\) is the homogeneity of this metric space. A metric space \((X, d)\) is **homogeneous** if, for any two points \(P, Q \in X\), there exists an isometry \(\varphi: X \to X\) such that \(\varphi(P) = Q\). In other words, a homogeneous metric space looks the same at every point, since any property of \((X, d)\) involving the point \(P\) also holds at any other point \(Q\), by translating this property through the isometry \(\varphi\) sending \(P\) to \(Q\).

Actually, the euclidean plane is not just homogeneous, it is **isotropic** in the sense that, for any two points \(P_1\) and \(P_2 \in \mathbb{R}^2\), and for any unit vectors \(\vec{v}_1\) at \(P_1\) and \(\vec{v}_2\) at \(P_2\), there is an isometry \(\varphi\) of \((\mathbb{R}^2, d_{\text{eucl}})\) which sends \(P_1\) to \(P_2\) and \(\vec{v}_1\) to \(\vec{v}_2\). We are here assuming that
the statement that \( \varphi \) sends the vector \( \vec{v}_1 \) to the vector \( \vec{v}_2 \) is intuitively clear; a more precise
definition, using the differential \( D_{P_1} \varphi \) of \( \varphi \) at \( P_1 \), will be given in Section 3.5.2.

As a consequence of the isotropy property, not only does the euclidean plane look the
same at every point, it also looks the same in every direction.

**Exercises for Chapter 2**

**Exercise 2.1.** Using the expression given in Equation (2.1) and a suitable parametrization of the line segment
\([P_0, P_1]\) going from \( P_0 = (x_0, y_0) \) to \( P_1 = (x_1, y_1) \), show that the euclidean length of \([P_0, P_1]\) is equal to
\[\ell_{\text{euc}}([P_0, P_1]) = \sqrt{(x_1 - x_0)^2 + (y_1 - y_0)^2}.\]

**Exercise 2.2.** The goal of this exercise is to rigorously prove that the line segment \([P, Q]\) is the shortest curve
going from \( P \) to \( Q \). Namely, consider a piecewise differentiable curve \( \gamma \) going from \( P \) to \( Q \). We want to show that
the euclidean length \( \ell_{\text{euc}}(\gamma) \) defined by Equation (2.1) is greater than or equal to the length \( \ell_{\text{euc}}([P, Q]) \) of the line segment \([P, Q]\).

- **a.** First consider the case where \( P = (x_0, y_0) \) and \( Q = (x_0, y_1) \) sit on the same vertical line of equation \( x = x_0 \).
  - Show that the euclidean length \( \ell_{\text{euc}}(\gamma) \) is greater than or equal to \( |y_1 - y_0| = \ell_{\text{euc}}([P, Q]) \).
- **b.** In the general case, let \( \varphi: (x, y) \mapsto (x \cos \theta - y \sin \theta, x \sin \theta + y \cos \theta) \) be a rotation such that \( \varphi(P) \) and \( \varphi(Q) \) sit
  - on the same vertical line. Show that the curve \( \varphi(\gamma) \), going from \( \varphi(P) \) to \( \varphi(Q) \), has the same euclidean length as \( \gamma \).
- **c.** Combine Parts a and b to conclude that \( \ell_{\text{euc}}(\gamma) \geq \ell_{\text{euc}}([P, Q]) \).

**Exercise 2.3.** Rigorously prove that the euclidean distance function \( d_{\text{euc}} \), as defined by Equation (2.2), is a
distance function on \( \mathbb{R}^2 \). You may need to use the result of Exercise 2.2 to show that \( d_{\text{euc}}(P, Q) = 0 \) only when
\( P = Q \). Note that the proof of the triangle inequality (for which you may find it useful to consult the proof of
Lemma 3.1 in the next chapter) is greatly simplified by our use of piecewise differentiable curves in the definition of
\( d_{\text{euc}} \).

**Exercise 2.4.** Let \((X, d)\) be a metric space.

- **a.** Show that \( d(P, Q) - d(P, Q') \leq d(Q, Q') \) for every \( P, Q, Q' \in X \).
- **b.** Conclude that \(|d(P, Q) - d(P, Q')| \leq d(Q, Q') \) for every \( P, Q, Q' \in X \).
- **c.** Use the above inequality to show that, for every \( P \in X \), the function \( d_P: X \to \mathbb{R} \) defined by \( d_P(Q) = d(P, Q) \)
is continuous, if we endow the real line \( \mathbb{R} \) with the usual metric for which the distance between \( a \) and \( b \in \mathbb{R} \) is
equal to the absolute value \( |a - b| \).

**Exercise 2.5.** Let \( \varphi: X \to X' \) be a map from the metric space \((X, d)\) to the metric space \((X', d')\). Show that \( \varphi \)
is continuous at \( P_0 \in X \) if and only if, for every \( \varepsilon > 0 \), there exists a \( \delta > 0 \) such that the image \( \varphi(B_d(P_0, \delta)) \) of the
ball \( B_d(P_0, \delta) \subset X \) is contained in the ball \( B_{d'}(\varphi(P_0), \varepsilon) \subset X' \).

**Exercise 2.6 (Product of metric spaces).** Let \((X, d)\) and \((X', d')\) be two metric spaces. On the product \( X \times X' =
\{(x, x'); x \in X, x' \in X'\} \) define
\[D: (X \times X') \times (X \times X') \to \mathbb{R}\]
by the property that \( D((x, x'), (y, y')) = \max\{d(x, y), d'(x', y')\} \) for every \((x, x'), (y, y') \in X \times X'\). Show that \( D \) is a
metric function on \( X \times X' \).

**Exercise 2.7.** On \( \mathbb{R}^2 = \mathbb{R} \times \mathbb{R} \), consider the metric function \( D \) provided by Exercise 2.6. Namely \( D((x, y), (x', y')) =
\max\{|x - x'|, |y - y'|\} \) for every \((x, y), (x', y') \in \mathbb{R}^2\).

- **a.** Show that \( \ell_{\text{euc}}(P, P') \leq d(P, P') \leq \ell_{\text{euc}}(P, P') \) for every \( P, P' \in \mathbb{R}^2 \).
- **b.** Let \( (P_n)_{n \in \mathbb{N}} \) be a sequence in \( \mathbb{R}^2 \). Show that \( (P_n)_{n \in \mathbb{N}} \) converges to a point \( P_\infty \in \mathbb{R}^2 \) for the metric \( D \) if and only if it converges to \( P_\infty \) for the metric \( \ell_{\text{euc}} \).
- **c.** Let \( \varphi: \mathbb{R}^2 \to X \) be a map from \( \mathbb{R}^2 \) to a metric space \((X, d)\). Show that \( \varphi \) is continuous for the metric \( D \) on \( \mathbb{R}^2 \) if and only if it is continuous for the metric \( \ell_{\text{euc}} \).

**Exercise 2.8 (Continuity and sequences).** Let \( \varphi: X \to X' \) be a map from the metric space \((X, d)\) to the metric space \((X', d')\).

- **a.** Suppose that \( \varphi \) is continuous at \( P_0 \). Show that, if \( P_1, P_2, \ldots, P_n, \ldots \) is a sequence which converges to \( P_0 \) in
  \((X, d)\), then \( \varphi(P_1), \varphi(P_2), \ldots, \varphi(P_n), \ldots \) is a sequence which converges to \( \varphi(P_0) \) in \((X', d')\).
- **b.** Suppose that \( \varphi \) is not continuous at \( P_0 \). Construct a number \( \varepsilon > 0 \) and a sequence \( P_1, P_2, \ldots, P_n, \ldots \) in \( X \)
such that \( d(P_n, P_0) < \frac{1}{n} \) and \( d(\varphi(P_n), \varphi(P_0)) \geq \varepsilon \) for every \( n \geq 1 \).
- **c.** Combine Parts a and b to show that \( \varphi \) is continuous at \( P_0 \) if and only if, for every sequence \( P_1, P_2, \ldots, P_n, \ldots \)
  converging to \( P_0 \) in \((X, d)\), the sequence \( \varphi(P_1), \varphi(P_2), \ldots, \varphi(P_n), \ldots \) converges to \( \varphi(P_0) \) in \((X', d')\).
Exercise 2.9. Let $d$ and $d'$ be two metrics on the same set $X$. Show that the identity map $I_d : (X, d) \to (X, d')$ is continuous if and only if every sequence $(P_n)_{n \in \mathbb{N}}$ which converges to some $P_\infty \in X$ for the metric $d$ also converges to $P_\infty$ for the metric $d'$. Possible hint: Compare Exercise 2.8.

Exercise 2.10. The euclidean metric of the euclidean plane is an example of a path metric, where the distance between two points $P$ and $Q$ is the infimum of the lengths of all curves joining $P$ to $Q$. In the plane $\mathbb{R}^2$, let $U$ be the U-shaped region enclosed by the polygonal curve with vertices $(0,0)$, $(0,2)$, $(1,2)$, $(1,1)$, $(2,1)$, $(2,2)$, $(3,2)$, $(3,0)$, $(0,0)$ occurring in this order. Endow $U$ with the metric $d_U$ defined by the property that $d_U(P, Q)$ is the infimum of the euclidean lengths of all piecewise differentiable curves joining $P$ to $Q$ and completely contained in $U$.

- a. Draw a picture of $U$.
- b. Show that $d_U(P, Q) \geq d_{euc}(P, Q)$ for every $P, Q \in U$.
- c. Show that $d_U$ is a metric function on $U$. It may be convenient to use Part b at some point of the proof.
- d. If $P_0$ is the point $(0,2)$, give a formula for the distance $d_U(P, P_0)$ in function of the coordinates of $P = (x, y)$.

This formula will involve several cases, according to where $P$ sits in $U$.

Exercise 2.11 (Lengths in metric spaces). In an arbitrary metric space $(X, d)$, the length $\ell_d(\gamma)$ of a curve $\gamma$ is defined as

$$\ell_d(\gamma) = \sup \left\{ \sum_{i=1}^{n} d(P_{i-1}, P_i); P_0, P_1, \ldots, P_n \in \gamma \text{ occur in this order along } \gamma \right\}.$$ 

In particular, the length may be infinite. For a differentiable curve $\gamma$ in the euclidean plane $(\mathbb{R}^2, d_{euc})$, we want to show that this length $\ell_{d_{euc}}(\gamma)$ coincides with the euclidean length $\ell_{euc}(\gamma)$ given by (2.1). For this, suppose that $\gamma$ is parametrized by the differentiable function $t \mapsto (x(t), y(t)), a \leq t \leq b$.

- a. Show that $\ell_{d_{euc}}(\gamma) \leq \ell_{euc}(\gamma)$.
- b. Cut the interval $[a, b]$ into $n$ intervals $[t_{i-1}, t_i]$ of length $\Delta t = (b - a)/n$. Set $P_i = \gamma(t_i) = \gamma(a + i\Delta t)$ for $i = 0, 1, \ldots, n$. Show that

$$d_{euc}(P_{i-1}, P_i) \geq \|\gamma'(t_{i-1})\| \Delta t - \frac{1}{2} K(\Delta t)^2$$

where $K = \max_{0 \leq t \leq b} \|\gamma''(t)\|$ denotes the maximum length of the second derivative vector $\gamma''(t) = (x''(t), y''(t))$ and where the length $\|\gamma'(t)\|$ of a vector $(u, v)$ is defined by the usual formula $\|\gamma'(t)\| = \sqrt{u^2 + v^2}$. You may need to use the Taylor formula from multivariable calculus, which says that for every $t, h$,

$$\gamma(t + h) = \gamma(t) + h\gamma'(t) + h^2 R_t(h)$$

where the remainder $R_t(h)$ is such that $\|R_t(h)\| \leq \frac{1}{2} K$.

- c. Use Part b to show that $\ell_{d_{euc}}(\gamma) \geq \ell_{euc}(\gamma)$.
- d. Combine Parts a and c to conclude that $\ell_{d_{euc}}(\gamma) = \ell_{euc}(\gamma)$.

Exercise 2.12. Consider the length $\ell_D(\gamma)$ of a curve $\gamma$ in a metric space $(X, d)$ defined as in Exercise 2.11, in the special case where $(X, d) = (\mathbb{R}^2, D)$ is the plane $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$ endowed with the product metric $D$ of Exercise 2.6, defined by the property that $D((x, y), (x', y')) = \max\{|x - x'|, |y - y'|\}$ for every $(x, y), (x', y') \in \mathbb{R}^2$.

- a. Show that $\ell_D(\gamma) \geq D(P, Q)$ for every curve $\gamma$ going from $P$ to $Q$.
- b. Show that the length $\ell_D([P, Q])$ of the line segment $[P, Q]$ is equal to $D(P, Q)$, so that $[P, Q]$ consequently has minimum length among all curves going from $P$ to $Q$.
- c. Give an example where there is another curve $\gamma$ going from $P$ to $Q$ which has minimum length $\ell_D(\gamma) = D(P, Q)$, and which is not the line segment $[P, Q]$.
- d. If $\gamma$ is differentiably parametrized by $t \mapsto (x(t), y(t)), a \leq t \leq b$, give a condition on the derivatives $x'(t)$ and $y'(t)$ which is equivalent to the property that $\gamma$ has minimum length over all curves going from $P = (x(a), y(a))$ to $Q = (x(b), y(b))$. (The answer depends on the relative position of $P$ and $Q$ with respect to each other.)
The hyperbolic plane

The hyperbolic plane is a metric space which is much less familiar than the euclidean plane that we discussed in the previous chapter. We introduce its basic properties, by proceeding in very close analogy with the euclidean plane.

3.1. The hyperbolic plane

The hyperbolic plane is the metric space consisting of the open half-plane

\[ \mathbb{H}^2 = \{(x, y) \in \mathbb{R}^2; y > 0\} = \{z \in \mathbb{C}; \text{Im}(z) > 0\} \]

endowed with a new metric \( d_{hyp} \) defined below. Recall that the \textit{imaginary part} \( \text{Im}(z) \) of a complex number \( z = x + iy \) is just the coordinate \( y \), while its \textit{real part} \( \text{Re}(z) \) is the coordinate \( x \).

To define the hyperbolic metric \( d_{hyp} \), we first define the \textit{hyperbolic length} of a curve \( \gamma \), parametrized by the differentiable vector-valued function

\[ t \mapsto (x(t), y(t)) \quad a \leq t \leq b, \]

as

\[ \ell_{hyp}(\gamma) = \int_{a}^{b} \sqrt{x'(t)^2 + y'(t)^2} \frac{dt}{y(t)} \]

Again, an application of the chain rule shows that this hyperbolic length is independent of the parametrization of \( \gamma \). The definition of the hyperbolic length also immediately extends to piecewise differentiable curves, by taking the sum of the hyperbolic length of the differentiable pieces, or by allowing finitely many jump discontinuities in the integrand of (3.1).

![Figure 3.1. The hyperbolic plane](image-url)
The hyperbolic distance between two points $P$ and $Q$ is the infimum of the hyperbolic lengths of all piecewise differentiable curves $\gamma$ going from $P$ to $Q$, namely

\begin{equation}
  d_{\text{hyp}}(P,Q) = \inf \{ \ell_{\text{hyp}}(\gamma); \gamma \text{ goes from } P \text{ to } Q \}
\end{equation}

Note the analogy with our definition of the euclidean distance in Chapter 2.

The hyperbolic distance $d_{\text{hyp}}$ is at first somewhat unintuitive. For instance, we will see in later sections that the hyperbolic distance between the points $P'$ and $Q'$ indicated in Figure 3.1 is the same as the hyperbolic distance from $P$ to $Q$. Also, among the curves joining $P$ to $Q$, the one with the shortest hyperbolic length is the circle arc represented. With practice, we will become more comfortable with the geometry of the hyperbolic plane, and see that it actually shares many important features with the euclidean plane.

But first, let us prove that the hyperbolic distance $d_{\text{hyp}}$ is really a distance function.

**Lemma 3.1.** The function

$$d_{\text{hyp}}: \mathbb{H}^2 \times \mathbb{H}^2 \rightarrow \mathbb{R}$$

defined by (3.2) is a distance function.

**Proof.** We have to check the four conditions in the definition of a distance function. The condition $d_{\text{hyp}}(P,Q) \geq 0$ is immediate, as is the symmetry condition $d_{\text{hyp}}(Q,P) = d_{\text{hyp}}(P,Q)$.

To prove the Triangle Inequality, consider three points $P$, $Q$, $R \in \mathbb{H}^2$. Pick an arbitrary $\varepsilon > 0$. By definition of the hyperbolic distance as an infimum of hyperbolic lengths, there exists a piecewise differentiable curve $\gamma$ going from $P$ to $Q$ such that $\ell_{\text{hyp}}(\gamma) < d_{\text{hyp}}(P,Q) + \frac{1}{2}\varepsilon$, and a piecewise differentiable curve $\gamma'$ going from $Q$ to $R$ such that $\ell_{\text{hyp}}(\gamma') < d_{\text{hyp}}(Q,R) + \frac{1}{2}\varepsilon$.

Chaining together these two curves $\gamma$ and $\gamma'$, one obtains a piecewise differentiable curve $\gamma''$ joining $P$ to $R$ whose length is

$$\ell_{\text{hyp}}(\gamma'') = \ell_{\text{hyp}}(\gamma) + \ell_{\text{hyp}}(\gamma') < d_{\text{hyp}}(P,Q) + d_{\text{hyp}}(Q,R) + \varepsilon.$$

As a consequence,

$$d_{\text{hyp}}(P,R) < d_{\text{hyp}}(P,Q) + d_{\text{hyp}}(Q,R) + \varepsilon.$$

Since this property holds for every $\varepsilon > 0$, we conclude that $d_{\text{hyp}}(P,R) \leq d_{\text{hyp}}(P,Q) + d_{\text{hyp}}(Q,R)$ as required.

Note that our use of piecewise differentiable curves, instead of just differentiable curves, greatly simplified this proof of the triangle inequality. (When $\gamma$ and $\gamma'$ are differentiable, the same is usually not true for $\gamma''$ since it may have a ‘corner’ at the junction of $\gamma$ and $\gamma'$).

The only condition which requires some serious thought is the fact that $d_{\text{hyp}}(P,Q) > 0$ if $P \neq Q$. Namely, we need to make sure that we cannot go from $P$ to $Q$ by curves whose hyperbolic lengths are arbitrary small.

Consider a piecewise differentiable curve $\gamma$ going from $P$ to $Q$, parametrized by the piecewise differentiable function

$$t \mapsto (x(t),y(t)) \quad a \leq t \leq b,$$ 

with $P = (x(a),y(a))$ and $Q = (x(b),y(b))$. We will split the argument in two cases.
If \( \gamma \) does not go too high, so that \( y(t) \leq 2y(a) \) for every \( t \in [a, b] \),

\[
\ell_{\text{hyp}}(\gamma) = \int_a^b \frac{\sqrt{x'(t)^2 + y'(t)^2}}{y(t)} \, dt
\]

\[
\geq \int_a^b \frac{\sqrt{x'(t)^2 + y'(t)^2}}{2y(a)} \, dt = \frac{1}{2y(a)}\ell_{\text{euc}}(\gamma)
\]

\[
\geq \frac{1}{2y(a)}d_{\text{euc}}(P, Q).
\]

Otherwise, \( \gamma \) crosses the horizontal line \( L \) of equation \( y = 2y(a) \). Let \( t_0 \) be the first value of \( t \) for which this happens; namely \( y(t_0) = 2y(a) \), and \( y(t) < 2y(a) \) for every \( t < t_0 \). Let \( \gamma' \) denote the part of \( \gamma \) corresponding to the values of \( t \) with \( a \leq t \leq t_0 \). This curve \( \gamma' \) joins \( P \) to the point \( (x(t_0), y(t_0)) \in L \), so that its euclidean length \( \ell_{\text{euc}}(\gamma') \) is greater than or equal to the euclidean distance from \( P \) to the line \( L \), which itself is equal to \( y(a) \). Therefore,

\[
\ell_{\text{hyp}}(\gamma) \geq \ell_{\text{hyp}}(\gamma') = \int_a^{t_0} \frac{\sqrt{x'(t)^2 + y'(t)^2}}{y(t)} \, dt
\]

\[
\geq \int_a^{t_0} \frac{\sqrt{x'(t)^2 + y'(t)^2}}{2y(a)} \, dt = \frac{1}{2y(a)}\ell_{\text{euc}}(\gamma')
\]

\[
\geq \frac{1}{2y(a)}y(a) = \frac{1}{2}.
\]

In both cases, we found that \( \ell_{\text{hyp}}(\gamma) \geq C \) for a positive constant

\[
C = \min \left\{ \frac{1}{2y(a)}d_{\text{euc}}(P, Q), \frac{1}{2} \right\} > 0
\]

which depends only on \( P \) and \( Q \) (remember that \( y(a) \) is the \( y \)-coordinate of \( P \)). If follows that \( d_{\text{hyp}}(P, Q) \geq C > 0 \) cannot be 0 if \( P \neq Q \). \( \square \)

### 3.2. Some isometries of the hyperbolic plane

The hyperbolic plane \((\mathbb{H}^2, d_{\text{hyp}})\) has many symmetries. Actually, we will see that it is as symmetric as the euclidean plane.

#### 3.2.1. Homotheties and horizontal translations.

Some of these isometries are surprising at first. These include the **homotheties** defined by \( \varphi(x, y) = (\lambda x, \lambda y) \) for some \( \lambda > 0 \). Indeed, if the piecewise differentiable curve \( \gamma \) is parametrized by

\[
t \mapsto (x(t), y(t)), \quad a \leq t \leq b,
\]

its image \( \varphi(\gamma) \) under \( \varphi \) is parametrized by

\[
t \mapsto (\lambda x(t), \lambda y(t)), \quad a \leq t \leq b.
\]

Therefore

\[
\ell_{\text{hyp}}(\varphi(\gamma)) = \int_a^b \frac{\sqrt{\lambda^2 x'(t)^2 + \lambda^2 y'(t)^2}}{\lambda y(t)} \, dt
\]

\[
= \int_a^b \frac{\sqrt{x'(t)^2 + y'(t)^2}}{y(t)} \, dt
\]

\[
= \ell_{\text{hyp}}(\gamma).
\]
Since \( \varphi \) establishes a one-to-one correspondence between curves joining \( P \) to \( Q \) and curves joining \( \varphi(P) \) to \( \varphi(Q) \), it follows from the definition of the hyperbolic metric that \( d_{hyp}(\varphi(P), \varphi(Q)) = d_{hyp}(P, Q) \) for every \( P, Q \in \mathbb{H}^2 \). This proves that the homothety \( \varphi \) is indeed an isometry of \((\mathbb{H}^2, d_{hyp})\).

The horizontal translations defined by \( \varphi(x, y) = (x + x_0, y) \) for some \( x_0 \in \mathbb{R} \) are more obvious isometries of \((\mathbb{H}^2, d_{hyp})\), as is the reflection \( \varphi(x, y) = (-x, y) \) across the \( y \)-axis.

### 3.2.2. The homogeneity property of the hyperbolic plane

The isometries obtained by composing homotheties and horizontal translations are enough to prove that the hyperbolic plane is homogeneous. Recall that the composition of two maps \( \varphi: X \to X' \) and \( \psi: X' \to X'' \) is the map \( \psi \circ \varphi: X \to X'' \) defined by \( \psi \circ \varphi(P) = \psi(\varphi(P)) \) for every \( P \in X \). If, in addition, \( X, X' \) and \( X'' \) are metric spaces, if \( \varphi \) is an isometry from \((X, d)\) to \((X', d')\), and if \( \psi \) is an isometry from \((X', d')\) to \((X'', d'')\), then \( \psi \circ \varphi \) is an isometry from \((X, d)\) to \((X'', d'')\) since

\[
d''(\psi \circ \varphi(P), \psi \circ \varphi(Q)) = d''(\psi(\varphi(P)), \psi(\varphi(Q)))
\]
\[
= d'(\varphi(P), \varphi(Q))
\]
\[
= d(P, Q).
\]

**Proposition 3.2.** The hyperbolic plane \((\mathbb{H}^2, d_{hyp})\) is homogeneous. Namely, for every \( P, Q \in \mathbb{H}^2 \), there exists an isometry \( \varphi \) of \((\mathbb{H}^2, d_{hyp})\) such that \( \varphi(P) = Q \).

**Proof.** If \( P = (a, b) \) and \( Q = (c, d) \in \mathbb{H}^2 \), with \( b, d > 0 \), the homothety \( \varphi \) of ratio \( \lambda = \frac{4}{b} \) sends \( P \) to the point \( R = (\frac{ad}{b}, d) \) with the same \( y \)-coordinate \( d \) as \( Q \). Then the horizontal translation \( \psi(x, y) = (x + c - \frac{ad}{b}, y) \) sends \( R \) to \( Q \). The composition \( \psi \circ \varphi \) now provides an isometry sending \( P \) to \( Q \). \( \square \)

### 3.2.3. The standard inversion

We now consider an even less obvious isometry of \((\mathbb{H}^2, d_{hyp})\). The standard inversion, or inversion across the unit circle, or inversion for short, is defined by

\[
\varphi(x, y) = \left( \frac{x}{x^2 + y^2}, \frac{y}{x^2 + y^2} \right).
\]

This map is better understood in polar coordinates, as it sends the point with polar coordinates \([r, \theta]\) to the point with polar coordinates \([\frac{1}{r}, \theta]\).

![Figure 3.2. The inversion across the unit circle](image)

**Lemma 3.3.** The inversion across the unit circle is an isometry of the hyperbolic plane \((\mathbb{H}^2, d_{hyp})\).
3.3. SHORTEST CURVES IN THE HYPERBOLIC PLANE

PROOF. If $\gamma$ is a piecewise differentiable curve parametrized by

$$t \mapsto (x(t), y(t)), \quad a \leq t \leq b,$$

its image $\varphi(\gamma)$ under the inversion $\varphi$ is parametrized by

$$t \mapsto (x_1(t), y_1(t)), \quad a \leq t \leq b$$

with

$$x_1(t) = \frac{x(t)}{x(t)^2 + y(t)^2} \quad \text{and} \quad y_1(t) = \frac{y(t)}{x(t)^2 + y(t)^2}$$

Then

$$x_1'(t) = \frac{(y(t)^2 - x(t)^2)x'(t) - 2x(t)y(t)y'(t)}{(x(t)^2 + y(t)^2)^2}$$

and

$$y_1'(t) = \frac{(x(t)^2 - y(t)^2)y'(t) - 2x(t)y(t)x'(t)}{(x(t)^2 + y(t)^2)^2}$$

so that, after simplifications,

$$x_1'(t)^2 + y_1'(t)^2 = \frac{x'(t)^2 + y'(t)^2}{(x(t)^2 + y(t)^2)^2}.$$ 

It follows that

$$\ell_{\text{hyp}}(\varphi(\gamma)) = \int_a^b \frac{\sqrt{x_1'(t)^2 + y_1'(t)^2}}{y_1(t)} dt$$

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

$$= \ell_{\text{hyp}}(\gamma).$$

As before, this shows that the inversion $\varphi$ is an isometry of the hyperbolic plane. \hfill \Box

3.3. Shortest curves in the hyperbolic plane

In euclidean geometry, the shortest curve joining two points is the line segment with these two points as end points. We want to identify the shortest curve between two points in the hyperbolic plane.

We begin with a special case.

**Lemma 3.4.** If $P_0 = (x_0, y_0), P_1 = (x_0, y_1) \in \mathbb{H}^2$ are located on the same vertical line, then the line segment $[P_0, P_1]$ has the shortest hyperbolic length among all piecewise differentiable curves going from $P_0$ to $P_1$. In addition, the hyperbolic length of any other curve joining $P_0$ to $P_1$ has strictly larger hyperbolic length, and

$$d_{\text{hyp}}(P_0, P_1) = \ell_{\text{hyp}}([P_0, P_1]) = \ln \left\lvert \frac{y_1}{y_0} \right\rvert.$$ 

**Proof.** Assuming $y_0 \leq y_1$ without loss of generality, let us first compute the hyperbolic length of $[P_0, P_1]$. Parametrize this line segment by

$$t \mapsto (x_0, t), \quad y_0 \leq t \leq y_1.$$
Then,

\[ \ell_{\text{hyp}}([P_0, P_1]) = \int_{y_0}^{y_1} \frac{\sqrt{0^2 + 1^2}}{t} dt = \ln \frac{y_1}{y_0}. \]

Now, consider a piecewise differentiable curve \( \gamma \) going from \( P_0 \) to \( P_1 \), and parametrized by

\[ t \mapsto (x(t), y(t)), \quad a \leq t \leq b. \]

Its hyperbolic length is

\[ \ell_{\text{hyp}}(\gamma) = \int_{a}^{b} \frac{\sqrt{x'(t)^2 + y'(t)^2}}{y(t)} dt \geq \int_{a}^{b} \frac{|y'(t)|}{y(t)} dt \]

\[ \geq \int_{a}^{b} \frac{y'(t)}{y(t)} dt = \ln \frac{y(b)}{y(a)} = \ln \frac{y_1}{y_0} = \ell_{\text{hyp}}([P_0, P_1]). \]

In addition, for the first term to be equal to the last one, the above two inequalities must be equalities. Equality in the first inequality requires that the function \( x(t) \) be constant, while equality in the second one implies that \( y(t) \) is weakly increasing. This shows that the curve \( \gamma \) is equal to the line segment \([P_0, P_1]\) if \( \ell_{\text{hyp}}(\gamma) = \ell_{\text{hyp}}([P_0, P_1]) \).

For future reference, we note the following estimate, which is proved by the same argument as the second half of the proof of Lemma 3.4.

**Lemma 3.5.** For any two points \( P_0 = (x_0, y_0), P_1 = (x_1, y_1) \in \mathbb{H}^2 \),

\[ d_{\text{hyp}}(P_0, P_1) \geq \ln \left| \frac{y_1}{y_0} \right|. \]

In our determination of shortest curves in the hyperbolic plane, the next step is the following.

**Lemma 3.6.** For any \( P, Q \in \mathbb{H}^2 \) which are not on the same vertical line, there exists an isometry of the hyperbolic plane \((\mathbb{H}^2, d_{\text{hyp}})\) such that \( \varphi(P) \) and \( \varphi(Q) \) are on the same vertical line. In addition, the line segment \([\varphi(P), \varphi(Q)]\) is the image under \( \varphi \) of the unique circle arc joining \( P \) to \( Q \) and centered on the \( x \)-axis.

**Proof.** Since \( P \) and \( Q \) are not on the same vertical line, the perpendicular bisector line of \( P \) and \( Q \) intersects the \( x \)-axis at some point \( R \). The point \( R \) is equidistant from \( P \) and \( Q \) for the euclidean metric, so that there is a circle \( C \) centered at \( R \) and passing through \( P \) and \( Q \). Note that \( C \) is the only circle passing through \( P \) and \( Q \) and centered on the \( x \)-axis.
The circle $C$ intersects the $x$–axis in two points. Let $\varphi_1$ be a horizontal translation sending one of these points to $(0, 0)$. Then $C' = \varphi_1(C)$ is a circle passing through the origin and centered at some point $(a, 0)$.

In particular, the equation of the circle $C'$ in polar coordinates is $r = 2a \cos \theta$. Its image under the inversion $\varphi_2$ is the curve of polar coordinate equation $r = \frac{1}{2a \cos \theta}$, namely the vertical line $L$ whose equation in cartesian coordinates is $x = \frac{1}{2a}$.

The composition $\varphi_2 \circ \varphi_1$ sends the circle $C$ to the vertical line $L$. In particular, it sends the points $P$ and $Q$ to two points on the vertical line $L$. Restricting $\varphi_2 \circ \varphi_1$ to points in $\mathbb{H}^2$ then provides the isometry $\varphi$ of $(\mathbb{H}^2, d_{hyp})$ that we were looking for.

Lemma 3.6 can immediately be extended to the case where $P$ and $Q$ sit on the same vertical line $L$, by interpreting $L$ as a circle of infinite radius whose center is located at infinity on the $x$–axis. Indeed, the vertical line $L$ of equation $x = a$ can be seen as the limit, as $x$ tends to $+\infty$ or to $-\infty$, of the circle of radius $|x - a|$ centered at the point $(x, 0)$. With this convention, any two $P, Q \in \mathbb{H}^2$ can be joined by a unique circle arc centered on the $x$–axis.

**Theorem 3.7.** Among all curves joining $P$ to $Q$ in $\mathbb{H}^2$, the circle arc centered on the $x$–axis (possibly a vertical line segment) is the unique one that is shortest for the hyperbolic length $\ell_{hyp}$.

**Proof.** If $P$ and $Q$ are on the same vertical line, this is Lemma 3.4.

Otherwise, Lemma 3.6 provides an isometry $\varphi$ sending $P$ and $Q$ to two points $P'$ and $Q'$ on the same vertical line $L$. By Lemma 3.4, the shortest curve from $P'$ to $Q'$ is the line segment $[P', Q']$. Since an isometry sends shortest curves to shortest curves, the shortest curve from $P$ to $Q$ is the image of the line segment $[P', Q']$ under the inverse isometry $\varphi^{-1}$. By the second statement of Lemma 3.6, this image is the circle arc joining $P$ to $Q$ and centered on the $x$–axis.

In a metric space where the distance function is defined by taking the infimum of the arc lengths of certain curves, such as the euclidean plane and the hyperbolic plane, there is a technical term for “shortest curve”. More precisely, a **geodesic** is a curve $\gamma$ such that, for every $P \in \gamma$ and for every $Q \in \gamma$ sufficiently close to $P$, the section of $\gamma$ joining $P$ to $Q$ is the shortest curve joining $P$ to $Q$ (for the arc length considered).

For instance, Proposition 2.1 says that geodesics in the euclidean plane $(\mathbb{R}^2, d_{euc})$ are line segments, whereas Theorem 3.7 shows that geodesics in the hyperbolic plane $(\mathbb{H}^2, d_{hyp})$ are circle arcs centered on the $x$–axis. By convention, line segments and circle arcs may include
some, all, or none of their end points (in much the same way as an interval in the number line $\mathbb{R}$ may be open, closed or semi-open).

A **complete geodesic** is a geodesic which cannot be extended to a larger geodesic. From the above observations, complete geodesics of the euclidean plane are straight lines. Complete geodesics of the hyperbolic plane are open semi-circles centered on the $x$–axis and delimited by two points of the $x$–axis (including vertical half-lines going from a point on the $x$–axis to infinity).

For future reference, we now prove the following technical result.

**Lemma 3.8.** Let $P_0 = (0, y_0)$ and $P_1 = (0, y_1)$ be two points of the upper half $L = \{(0, y); y > 0\} \subset \mathbb{H}^2$ of the $y$–axis, with $y_1 > y_0$, and let $g$ be a complete hyperbolic geodesic passing through $P_0$. See Figure 3.5. Then the following are equivalent:

1. $P_0$ is the point of $g$ that is closest to $P_1$ for the hyperbolic distance $d_{\text{hyp}}$;
2. $g$ is the complete geodesic $g_0$ that is orthogonal to $L$ at $P_0$, namely is the euclidean semi-circle of radius $y_0$ centered at $(0, 0)$ and joining $(y_0, 0)$ to $(-y_0, 0)$.

**Figure 3.5**

**Proof.** Lemmas 3.4 and 3.5 show that, for every point $P = (u, v)$ on the geodesic $g_0$,

$$d_{\text{hyp}}(P_1, P) \geq \ln \frac{y_1}{v} \geq \ln \frac{y_1}{y_0} = d_{\text{hyp}}(P_1, P_0).$$

As a consequence, the point $P_0$ is closest to $P_1$ among all points of $g_0$.

Conversely, if $g$ is another complete hyperbolic geodesic which passes through $P_0$ and makes an angle of $\theta \neq \frac{\pi}{2}$ with $L$ at $P_0$, we want to find a point $P \in g$ with $d_{\text{hyp}}(P_1, P) < d_{\text{hyp}}(P_1, P_0)$.

For $P = (u, v) \in g$, the standard parametrization of the line segment $[P_1, P]$ gives that its hyperbolic length is equal to

$$\ell_{\text{hyp}}([P_1, P]) = \int_0^1 \sqrt{\frac{u^2 + (v - y_1)^2}{y_1 + t(v - y_1)}} \, dt = \sqrt{\frac{u^2 + (v - y_1)^2}{y_1 - v}} \ln \frac{y_1}{v}.$$

We now let the point $P = (u, v)$ vary on the geodesic $g$ near $P_0$. When $u = 0$, $v = y_0$ and $\frac{dv}{du} = \cot \theta$. Differentiating the above formula then gives that, still at $u = 0$,

$$\frac{d}{du} \ell_{\text{hyp}}([P_1, P]) = -\frac{1}{y_0} \cot \theta.$$
In particular, unless $\theta = \frac{\pi}{2}$, this derivative is different from 0 and there exists near $P_0 = (0, y_0)$ a point $P = (u, v)$ of $g$ such that
\[
d_{\text{hyp}}(P_1, P) \leq \ell_{\text{hyp}}([P_1, P]) < \ell_{\text{hyp}}([P_1, P_0]) = d_{\text{hyp}}([P_1, P_0]). \quad \Box
\]

### 3.4. All isometries of the hyperbolic plane

So far, we have encountered three types of isometries of the hyperbolic plane: homotheties, horizontal translations and the inversion. In this section, we describe all of its isometries.

It is convenient to use complex numbers. In this framework,
\[
\mathbb{H}^2 = \{ z \in \mathbb{C}; \Im(z) > 0 \}
\]
where the imaginary part $\Im(z)$ is the $y$–coordinate of $z = x + iy$.

In complex coordinates, a horizontal translation is of the form $z \mapsto z + x_0$ with $x_0 \in \mathbb{R}$, a homothety is of the form $z \mapsto \lambda z$ for a real number $\lambda > 0$, and the inversion is of the form $z \mapsto \frac{\bar{z}}{|z|^2} = \frac{1}{z}$, where $\bar{z} = x - iy$ is the complex conjugate of $z = x + iy$ and where $|z| = \sqrt{x^2 + y^2} = \sqrt{zz}$ is its modulus.

We can obtain more examples of isometries by composition of these ones. Recall that the composition $\psi \circ \varphi$ of two maps $\varphi$ and $\psi$ is defined by $\psi \circ \varphi(P) = \psi(\varphi(P))$, and that the composition of two isometries is an isometry.

**Lemma 3.9.** All maps of the form
\[
\begin{align*}
(3.3) & \quad z \mapsto \frac{az + b}{cz + d} \quad \text{with } a, b, c, d \in \mathbb{R} \text{ and } ad - bc = 1 \\
(3.4) & \quad z \mapsto \frac{c\bar{z} + d}{a\bar{z} + b} \quad \text{with } a, b, c, d \in \mathbb{R} \text{ and } ad - bc = 1
\end{align*}
\]
are isometries of the hyperbolic plane $(\mathbb{H}^2, d_{\text{hyp}})$.

**Proof.** We will show that every such map is a composition of horizontal translations $z \mapsto z + x_0$ with $x_0 \in \mathbb{R}$, of homotheties $z \mapsto \lambda z$ with $\lambda > 0$, and of inversions $z \mapsto \frac{1}{z}$. Since a composition of isometries is an isometry, this will prove the result.

When $a \neq 0$, the map of Equation (3.4) is the composition of
\[
\begin{align*}
z & \mapsto z + \frac{b}{a} , \quad z \mapsto \frac{1}{\bar{z}} , \quad z \mapsto \frac{1}{a\bar{z}} \bar{z} \quad \text{and} \quad z \mapsto z + \frac{c}{a}.
\end{align*}
\]
In particular, this map is the composition of several isometries of $\mathbb{H}^2$, and is therefore an isometry of $\mathbb{H}^2$.

Composing once more with $z \mapsto \frac{1}{\bar{z}}$, we obtain the map of Equation (3.3), thereby showing that this map is also an isometry of the hyperbolic plane when $a \neq 0$.

When $a = 0$, so that $c \neq 0$, the map of Equation (3.3) is the composition of
\[
z \mapsto \frac{cz + b + d}{cz + d},
\]
which is an isometry of $\mathbb{H}^2$ by the previous case, and of the horizontal translation $z \mapsto z - 1$. It follows that the map of Equation (3.3) is also an isometry of $\mathbb{H}^2$ when $a = 0$.

Finally, composing with $z \mapsto \frac{1}{\bar{z}}$ shows that the map of Equation (3.4) is an isometry of $\mathbb{H}^2$ when $a = 0$. \quad \Box
Conversely, we will show that every isometry of the hyperbolic plane is of one the two types considered in Lemma 3.9. The proof of this fact hinges on the following property.

**Lemma 3.10.** Let \( \varphi \) be an isometry of the hyperbolic plane \((\mathbb{H}^2, d_{hyp})\) such that \( \varphi(iy) = iy \) for every \( y > 0 \). Then either \( \varphi(z) = z \) for every \( z \in \mathbb{H}^2 \), or \( \varphi(z) = -\bar{z} \) for every \( z \).

**Proof.** Let \( L = \{iy; y > 0\} \) be the upper half of the \( y \)-axis. By hypothesis, \( \varphi \) fixes every point of \( L \).

For every \( iy \in L \), let \( g_y \) be the unique hyperbolic complete geodesic that passes through \( iy \) and is orthogonal to \( L \). Namely, \( g_y \) is the euclidean semi-circle of radius \( y \) centered at 0 and contained in \( \mathbb{H}^2 \). Since \( \varphi \) is an isometry and \( \varphi(iy) = iy \), we know that it sends \( g_y \) to a complete geodesic \( g \) passing through \( iy \). We will use Lemma 3.8 to prove that \( g = g_y \).

Indeed, this statement characterizes the geodesic \( g_y \) by the property that, for any \( y_1 > y \), the point \( iy \) is the point of \( g_y \) that is closest to \( iy_1 \). As a consequence, since \( \varphi \) is an isometry, \( \varphi(iy) = iy \) is the point of \( \varphi(g_y) = g \) that is closest to \( \varphi(iy_1) = iy_1 \). Lemma 3.8 then shows that \( g = g_y \), so that \( \varphi(g_y) = g_y \).

Now, if \( P = u + iv \) is a point of \( g_y \), its image \( \varphi(g_y) \) is one of the two points of \( g_y \) that are at distance \( d_{hyp}(P, iy) \) from \( iy \). One of these two points is \( P \), the other one is \( -u + iv \) by symmetry.

We conclude that \( \varphi(u + iv) = u + iv \) or \( -u + iv \) for every \( u + iv \in \mathbb{H}^2 \) (since \( u + iv \) belongs to some geodesic \( g_y \)). Since \( \varphi \) is an isometry, it is continuous. It follows that, either \( \varphi(u + iv) = u + iv \) for every \( u + iv \in \mathbb{H}^2 \), or \( \varphi(u + iv) = -u + iv \) for every \( u + iv \in \mathbb{H}^2 \). This can be rephrased as: Either \( \varphi(z) = z \) for every \( z \in \mathbb{H}^2 \), or \( \varphi(z) = -\bar{z} \) for every \( z \in \mathbb{H}^2 \). \( \square \)

A minor corollary of Lemma 3.9 is that \( \varphi(z) = \frac{az + b}{cz + d} \) with \( a, b, c, d \in \mathbb{R} \) and \( ad - bc = 1 \) sends the upper-half space \( \mathbb{H}^2 \) to itself; this can also be easily checked “by hand”. This map is not defined at the boundary point \( z = -\frac{d}{c} \). However, if we introduce a point \( \infty \) at infinity of the real line \( \mathbb{R} \) (without distinguishing between \( +\infty \) and \( -\infty \)) the same formula defines a map

\[
\varphi: \mathbb{R} \cup \{\infty\} \to \mathbb{R} \cup \{\infty\}
\]

by setting \( \varphi(-\frac{d}{c}) = \infty \) and \( \varphi(\infty) = \frac{a}{c} \). This map is specially designed to be continuous. Indeed,

\[
\lim_{x \to -\frac{d}{c}} \varphi(x) = \infty \quad \text{and} \quad \lim_{x \to \infty} \varphi(x) = \frac{a}{c}
\]

in the “obvious” sense made precise in Section 1.4 of the tool kit appendix.

The same applies to a map of the form \( \varphi(z) = \frac{cz + d}{az + b} \) with \( a, b, c, d \in \mathbb{R} \) and \( ad - bc = 1 \). These extensions are often convenient, as in the proof of the following statement.

**Theorem 3.11.** The isometries of the hyperbolic plane \((\mathbb{H}^2, d_{hyp})\) are exactly the maps of the form

\[
\varphi(z) = \frac{az + b}{cz + d} \quad \text{with} \quad a, b, c, d \in \mathbb{R} \quad \text{and} \quad ad - bc = 1
\]

or

\[
\varphi(z) = \frac{cz + d}{az + b} \quad \text{with} \quad a, b, c, d \in \mathbb{R} \quad \text{and} \quad ad - bc = 1
\]
3.5. LINEAR AND ANTLINERAL FRACTIONAL MAPS

Proof. We already proved in Lemma 3.9 that all maps of these two types are isometries of the hyperbolic plane.

Conversely, let \( \varphi \) be an isometry of \( \mathbb{H}^2 \) and consider again the positive part \( L = \{ iy; y > 0 \} \) of the \( y \)-axis. Since \( L \) is a complete geodesic of \( \mathbb{H}^2 \), its image under the isometry \( \varphi \) is also a complete geodesic of \( \mathbb{H}^2 \), namely a euclidean semi-circle bounded by two distinct points \( u, v \in \mathbb{R} \cup \{ \infty \} \). Here \( u \) or \( v \) will be \( \infty \) exactly when \( \varphi(L) \) is a vertical half-line. In addition, if we orient \( L \) from 0 to \( \infty \), we require without loss of generality that the corresponding orientation of \( \varphi(L) \) goes from \( u \) to \( v \).

First consider the case where \( u \) and \( v \) are both different from \( \infty \). The hyperbolic isometry

\[
\psi(z) = \frac{az - au}{cz - cv},
\]

with \( a \) and \( c \in \mathbb{R} \) chosen so that \( ac(u - v) = 1 \), sends \( u \) to 0 and \( v \) to \( \infty \). It follows that the composition \( \psi \circ \varphi \) fixes the two points 0 and \( \infty \). As a consequence, the isometry \( \psi \circ \varphi \) sends the complete geodesic \( L \) to itself, and respects its orientation. In particular, \( \psi \circ \varphi(i) = it \) for some \( t > 0 \). Replacing \( a \) by \( a/\sqrt{t} \) and \( c \) by \( c\sqrt{t} \) in the definition of \( \psi \), we can arrange that \( \psi \circ \varphi(i) = i \). Then, \( \psi \circ \varphi \) sends each \( iy \in L \) to a point of \( L \) which is at the same hyperbolic distance from \( i \) as \( iy \); since \( \psi \circ \varphi \) respects the orientation of \( L \), the only possibility is that \( \psi \circ \varphi(iy) = iy \) for every \( y > 0 \).

Applying Lemma 3.10, we conclude that, either \( \psi \circ \varphi(z) = z \) for every \( z \), or \( \psi \circ \varphi(z) = -\bar{z} \) for every \( z \). In the first case,

\[
\varphi(z) = \psi^{-1}(z) = \frac{-cvz + au}{cz + a}
\]

where the formula for the inverse function \( \psi^{-1} \) is obtained by solving the equation \( \psi(z') = z \) (compare Exercise 3.10). In the second case,

\[
\varphi(z) = \psi^{-1}(-\bar{z}) = \frac{cv\bar{z} + au}{c\bar{z} + a}
\]

In both cases, \( \varphi(z) \) is of the type requested.

It remains to consider the cases where \( u \) or \( v \) is \( \infty \). The argument is identical, using the isometries

\[
\psi(z) = \frac{-a}{cz - cv},
\]

with \( ac = 1 \) when \( u = \infty \), and

\[
\psi(z) = \frac{az - au}{c},
\]

with \( ac = 1 \) when \( v = \infty \).

The isometries of \( (\mathbb{H}^2, d_{hyp}) \) of the form \( \varphi(z) = \frac{az + b}{cz + d} \) with \( a, b, c, d \in \mathbb{R} \) and \( ad - bc = 1 \) are called linear fractional maps with real coefficients. Those of the form \( \varphi(z) = \frac{cz + d}{az + b} \) with \( a, b, c, d \in \mathbb{R} \) and \( ad - bc = 1 \) are antilinear fractional maps.

3.5. Linear and antilinear fractional maps

We establish in this section a few fundamental properties of linear and antilinear fractional maps. Since we will later need to consider maps of this type with arbitrary complex (and not just real) coefficients, we prove these properties in this higher level of generality.
In this context, a linear fractional map is a non-constant map \( \varphi \) of the form \( \varphi(z) = \frac{az + b}{cz + d} \) with complex coefficients \( a, b, c, d \in \mathbb{C} \). Elementary algebra shows that \( \varphi \) is non-constant exactly when \( ad - bc \neq 0 \). Dividing all coefficients by one of the two complex square roots \( \pm \sqrt{ad - bc} \), we can consequently arrange that \( ad - bc = 1 \) without changing the map \( \varphi \). We will systematically require that the coefficients \( a, b, c, d \in \mathbb{C} \) satisfy this condition \( ad - bc = 1 \).

So far, the map \( \varphi \) is not defined at \( z = \frac{d}{c} \). However, this can easily be fixed by introducing a point \( \infty \) at infinity of \( \mathbb{C} \). Let the Riemann sphere be the union \( \hat{\mathbb{C}} = \mathbb{C} \cup \{ \infty \} \).

Then a linear fractional \( \varphi(z) = \frac{az + b}{cz + d} \) with \( a, b, c, d \in \mathbb{C} \) and \( ad - bc = 1 \) defines a map \( \varphi: \hat{\mathbb{C}} \to \hat{\mathbb{C}} \) by setting \( \varphi(-\frac{b}{a}) = \infty \) and \( \varphi(\infty) = \frac{a}{c} \). This map is continuous for the obvious definition of continuity at infinity, because

\[
\lim_{z \to -\frac{b}{a}} \varphi(z) = \infty \quad \text{and} \quad \lim_{z \to \infty} \varphi(z) = \frac{a}{c}
\]

where these limits involving infinity are defined exactly as in Section 1.4 of the tool kit appendix, but replacing absolute values of real numbers by moduli (= absolute values) of complex numbers.

Similarly, a general antilinear fractional map is a map \( \varphi: \hat{\mathbb{C}} \to \hat{\mathbb{C}} \) of the form \( \varphi(z) = \frac{cz + d}{az + b} \) with \( a, b, c, d \in \mathbb{C} \) and \( ad - bc = 1 \), with the convention that \( \varphi(-\frac{b}{a}) = \infty \) and \( \varphi(\infty) = \frac{a}{c} \).

See Exercise 3.8 for an explanation of why the Riemann sphere \( \hat{\mathbb{C}} = \mathbb{C} \cup \{ \infty \} \) can indeed be considered as a sphere. See also Exercise 3.12 for another interpretation of \( \hat{\mathbb{C}} \), which sheds a different light on linear fractional maps.

### 3.5.1. Some special (anti)linear fractional maps.

We already encountered the homotheties

\[
z \mapsto \lambda z = \frac{\lambda \bar{z} z + 0}{0z + \lambda^{-\frac{1}{2}}} \quad \text{with positive real ratio } \lambda > 0.
\]

If we allow complex coefficients, we can also consider the rotations

\[
z \mapsto e^{i\theta} z = \frac{e^{i\theta} z + 0}{0z + e^{-i\frac{\theta}{2}}} \quad \text{of angle } \theta \in \mathbb{R} \text{ around the origin},
\]

and the translations

\[
z \mapsto z + z_0 = \frac{z + z_0}{0z + 1} \quad \text{for arbitrary complex numbers } z_0 \in \mathbb{C}.
\]

We also considered the inversion across the unit circle

\[
z \mapsto \frac{z}{|z|^2} = \frac{1}{\bar{z}} = \frac{0\bar{z} + 1}{\bar{z} + 0}.
\]

**Lemma 3.12.** Every linear or antilinear fractional map \( \varphi: \hat{\mathbb{C}} \to \hat{\mathbb{C}} \) is a composition of homotheties, translations, rotations, and inversions across the unit circle.
**Proof.** The proof is identical to the purely algebraic argument that we already used in the proof of Lemma 3.9 for linear and linear fractional maps with real coefficients. □

Actually, there is no reason to prefer the unit circle to any other circle. If $C$ is the circle of radius $R$ centered at the point $z_0 \in \mathbb{C}$, the **inversion** across the circle $C$ is the antilinear fractional map $\varphi$ defined by the property that

$$\varphi(z) - z_0 = R^2 \frac{z - z_0}{|z - z_0|^2}$$

or, equivalently, that

$$\varphi(z) = \frac{R^2 z + R^2 - |z_0|^2}{R \bar{z} - \frac{2z_0}{R}}.$$  

Namely, $\varphi$ sends $z$ to the point that is on the same ray issued from $z_0$ as $z$, and is at euclidean distance $R^2/d_{\text{euc}}(z, z_0)$ from $z_0$. This inversion fixes every point of the circle $C$, and exchanges the inside and the outside of $C$.

There is an interesting limit case of inversions as we let the center and the radius of the circle go to infinity. For given $t$, $t_0$ and $\theta_0 \in \mathbb{R}$, set $z_0 = te^{i\theta_0}$ and $R = t - t_0$. If we let $t$ tend to $+\infty$, the circle $C$ converges to the line $L$ that passes through the point $t_0 e^{i\theta_0}$ and makes an angle of $\theta_0 + \frac{\pi}{2}$ with the $x$-axis. On the other hand, the inversion $\varphi$ across $C$ converges to the map $z \mapsto -e^{2i\theta_0} \bar{z} + 2t_0 e^{i\theta_0}$, which is just the reflection across the line $L$.

In this way, if we interpret the line $L$ as a circle of infinite radius centered at infinity, we can also consider the euclidean reflection across $L$ as an inversion across this circle. Note that every line $L$ can be obtained in this way.

**3.5.2. Differentials.** Recall that, if $\varphi: U \to \mathbb{R}^2$ is a differentiable function defined on a region $U \subset \mathbb{R}^2$ by $\varphi(x, y) = (f(x, y), g(x, y))$, the **differential** or **tangent map** of $\varphi$ at a point $P_0 = (x_0, y_0)$ in the interior of $U$ is the linear map $D_{P_0} \varphi: \mathbb{R}^2 \to \mathbb{R}^2$ with matrix

$$
\begin{pmatrix}
\frac{\partial f}{\partial x}(P_0) & \frac{\partial f}{\partial y}(P_0) \\
\frac{\partial g}{\partial x}(P_0) & \frac{\partial g}{\partial y}(P_0)
\end{pmatrix}.
\end{pmatrix}
$$

Namely

$$D_{P_0} \varphi(\vec{v}) = D_{P_0} \varphi(a, b) = \left(a \frac{\partial f}{\partial x}(P_0) + b \frac{\partial f}{\partial y}(P_0), a \frac{\partial g}{\partial x}(P_0) + b \frac{\partial g}{\partial y}(P_0)\right)$$

for every vector $\vec{v} = (a, b) \in \mathbb{R}^2$.

The differential map $D_{P_0} \varphi$ also has the following geometric interpretation.

**Lemma 3.13.** Let the differentiable map $\varphi: U \to \mathbb{R}^2$ be defined over a region $U$ containing the point $P_0$ in its interior. Then, for every parametrized curve $\gamma$ in $U$ which passes through $P_0$ and is tangent to the vector $\vec{v}$ there, its image under $\varphi$ is tangent to the vector $D_{P_0} \varphi(\vec{v})$ at the point $\varphi(P_0)$.

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{figure3.6.png}
\caption{The geometry of the differential map}
\end{figure}
Proof. Suppose that the map \( \varphi \) is given by \( \varphi(x,y) = (f(x,y), g(x,y)) \), and that the curve \( \gamma \) is parametrized by \( t \mapsto (x(t), y(t)) \). If the point \( P_0 \) corresponds to \( t = t_0 \), namely if \( P_0 = (x(t_0), y(t_0)) \), then \( \vec{v} = (x'(t_0), y'(t_0)) \).

The image of the curve \( \gamma \) under \( \varphi \) is parametrized by \( t \mapsto \varphi(x(t), y(t)) = (f(x(t), y(t)), g(x(t), y(t))) \).

Applying the chain rule for functions of several variables, its tangent vector at \( \varphi(P_0) \) is equal to

\[
\frac{d}{dt} \left( f(x(t), y(t)), g(x(t), y(t)) \right)_{t=t_0} = \left( \frac{df}{dx}(P_0) x'(t_0) + \frac{df}{dy}(P_0) y'(t_0), \frac{dg}{dx}(P_0) x'(t_0) + \frac{dg}{dy}(P_0) y'(t_0) \right) = D_{P_0} \varphi(\vec{v}).
\]

An immediate consequence of this geometric interpretation is the following property.

**Corollary 3.14.**

\[ D_{P_0} (\psi \circ \varphi) = (D_{\varphi(P_0)} \psi) \circ (D_{P_0} \varphi). \]

The differential maps of linear and antilinear fractional maps have a particularly nice expression in complex coordinates.

**Proposition 3.15.** If the linear fractional map \( \varphi \) is defined by \( \varphi(z) = \frac{az + b}{cz + d} \) where \( a, b, c, d \in \mathbb{C} \) with \( ad - bc = 1 \), its differential map \( D_{z_0} \varphi : \mathbb{C} \to \mathbb{C} \) at \( z_0 \in \mathbb{C} \) with \( z_0 \neq -\frac{d}{c} \) is such that

\[ D_{z_0} \varphi(v) = \frac{1}{(cz_0 + d)^2} v \]

for every \( v \in \mathbb{C} \).

If the antilinear fractional map \( \psi \) is defined by \( \psi(z) = \frac{c\bar{z} + d}{a\bar{z} + b} \) where \( a, b, c, d \in \mathbb{C} \) with \( ad - bc = 1 \), its differential map \( D_{z_0} \psi : \mathbb{C} \to \mathbb{C} \) at \( z_0 \in \mathbb{C} \) with \( z_0 \neq -\frac{b}{a} \) is such that

\[ D_{z_0} \psi(v) = \frac{1}{(a\bar{z}_0 + b)^2} \bar{v} \]

for every \( v \in \mathbb{C} \).

**Proof.** We will use Lemma 3.13. Given \( v \in \mathbb{C} \) interpreted as a vector, consider the line segment \( \gamma \) parametrized by \( t \mapsto z(t) = z_0 + tv \). Note that \( z(0) = z_0 \) and that \( z'(0) = v \).
Lemma 3.13 then implies that
\[ D_{z_0} \varphi(v) = \frac{d}{dt} \varphi(z(t))_{|t=0} = \lim_{h \to 0} \frac{1}{h} \left( \varphi(z(h)) - \varphi(z(0)) \right) \]
\[ = \lim_{h \to 0} \frac{1}{h} \left( \frac{az_0 + ahv + b}{cz_0 + chv + d} - \frac{az_0 + b}{cz_0 + d} \right) v \]
\[ = \lim_{h \to 0} \frac{1}{(cz_0 + d)^2} v, \]
using the property that \( ad - bc = 1 \).

The argument is identical for the antilinear fractional map \( \psi \). □

For future reference, we note that the same computation yields:

**Complement 3.16.** If \( \varphi(z) = \frac{az + b}{cz + d} \) where \( ad - bc \) is not necessarily equal to 1, then

\[ D_{z_0} \varphi(v) = \frac{ad - bc}{(cz_0 + d)^2} v. \] □

A consequence of Proposition 3.15 is that the differential map of a linear fractional map is the composition of a homothety with a rotation, and the differential map of an antilinear fractional map is the composition of a homothety with a reflection. This has the following important consequence.

**Corollary 3.17.** The differential map \( D_{z_0} \varphi \) of a linear fractional map \( \varphi \) respects angles and orientation in the sense that, for any two non-zero vectors \( \vec{v}_1, \vec{v}_2 \in \mathbb{C} \), the oriented angle from \( D_{z_0} \varphi(\vec{v}_1) \) to \( D_{z_0} \varphi(\vec{v}_2) \) is the same as the oriented angle from \( \vec{v}_1 \) to \( \vec{v}_2 \), measuring oriented angles counterclockwise in \( \mathbb{C} \).

The differential map \( D_{z_0} \psi \) of an antilinear fractional linear map \( \psi \) respects angles and reverses orientation in the sense that, for any two non-zero vectors \( \vec{v}_1, \vec{v}_2 \in \mathbb{C} \), the oriented angle from \( D_{z_0} \psi(\vec{v}_1) \) to \( D_{z_0} \psi(\vec{v}_2) \) is the opposite of the oriented angle from \( \vec{v}_1 \) to \( \vec{v}_2 \). □

Incidentally, Corollary 3.17 shows that a linear fractional map cannot coincide with an antilinear fractional map.

### 3.5.3. (Anti)linear fractional maps and circles

Another fundamental property of linear and antilinear fractional maps is that they send circles to circles. For this, we have to include all lines as circles of infinite radius centered at infinity. More precisely, let a **circle** in the Riemann sphere \( \hat{\mathbb{C}} = \mathbb{C} \cup \{\infty\} \) be, either a euclidean circle in \( \mathbb{C} \), or the union \( L \cup \{\infty\} \) of a line \( L \subset \mathbb{C} \) and of the point \( \infty \).

**Proposition 3.18.** A linear or antilinear fractional map \( \varphi: \hat{\mathbb{C}} \to \hat{\mathbb{C}} \) sends each circle of \( \hat{\mathbb{C}} \) to a circle of \( \hat{\mathbb{C}} \).

**Proof.** By Lemma 3.12, \( \varphi \) is a composition of homotheties, rotations, translations and inversions across the unit circle. Since homotheties, rotations and translations clearly send circles to circles, it suffices to consider the case where \( \varphi \) is the inversion across the unit circle.

It is convenient to use polar coordinates. In polar coordinates \( r \) and \( \theta \), the circle \( C \) of radius \( R \) centered at \( z_0 = r_0 e^{i\theta_0} \) has equation

\[ r^2 - 2r r_0 \cos(\theta - \theta_0) + r_0^2 - R^2 = 0. \]
The inversion $\varphi$ sends the point with polar coordinates $[r, \theta]$ to the point of coordinates $\left[\frac{1}{r}, \theta\right]$. The image of the circle $C$ under $\varphi$ is therefore the curve of equation

$$\frac{1}{r^2} - \frac{2r_0}{r} \cos(\theta - \theta_0) + r_0^2 - R^2 = 0.$$ 

If $|r_0| \neq R$ or, equivalently, if the circle $C$ does not contain the origin 0, simplifying the above equation shows that this curve is the circle of radius $\frac{R}{|r_0^2 - R^2|}$ centered at $\frac{z_0}{r_0^2 - R^2}$.

If $|z_0| = R$, we get the curve of polar equation $r = \frac{2r_0 \cos(\theta - \theta_0)}{1}$, which of course is a line.

Finally, we need to consider the case where $C$ is a line. In polar coordinates, its equation is of the form $r = \frac{1}{2r_0 \cos(\theta - \theta_0)}$ for some $r_0$ and $\theta_0$. Then its image under $\varphi$ has equation $r = 2r_0 \cos(\theta - \theta_0)$, and consequently is a circle passing through the origin. □

### 3.6. The hyperbolic norm

If $\vec{v} = (a, b)$ is a vector in $\mathbb{R}^2$, its **euclidean magnitude** or **euclidean norm** is its usual length

$$||\vec{v}||_{\text{euc}} = \sqrt{a^2 + b^2}.$$ 

For instance, if $\vec{v}$ is the velocity of a particle moving in the euclidean plane, $||\vec{v}||_{\text{euc}}$ describes the speed of this particle.

In the hyperbolic plane, distances are measured differently according to where we are in the plane, and consequently so are speeds. If $\vec{v}$ is a vector based at the point $z \in \mathbb{H}^2 \subset \mathbb{C}$, its **hyperbolic norm** is

$$||\vec{v}||_{\text{hyp}} = \frac{1}{\text{Im}(z)} ||\vec{v}||_{\text{euc}}.$$ 

To justify this definition, let $\gamma$ be a curve in $\mathbb{H}^2$, parametrized by $t \mapsto z(t)$, $a \leq t \leq b$. In particular, the tangent vector of $\gamma$ at the point $z(t)$ is the derivative $z'(t)$, and must be considered as a vector based at $z(t)$. Then, the euclidean and hyperbolic lengths of $\gamma$ are given by the very similar formulas

$$\ell_{\text{euc}}(\gamma) = \int_a^b ||z'(t)||_{\text{euc}} \, dt$$

and

$$\ell_{\text{hyp}}(\gamma) = \int_a^b ||z'(t)||_{\text{hyp}} \, dt.$$ 

If $\varphi$ is a differentiable map and $\vec{v}$ is a vector based at $P$, its image $D_P \varphi(\vec{v})$ under the differential map is a vector based at $\varphi(P)$. Indeed, see the geometric interpretation of the differential $D_P \varphi$ given by Lemma 3.13.

**Lemma 3.19.** If $\varphi$ be an isometry of $(\mathbb{H}^2, d_{\text{hyp}})$, then $||D_{z_0} \varphi(\vec{v})||_{\text{hyp}} = ||\vec{v}||_{\text{hyp}}$ for every vector $\vec{v}$ based at $z_0 \in \mathbb{H}^2$.

**Proof.** We could go back to basic principles about the metric $d_{\text{hyp}}$, but it is easier to use a straight computation.
Consider the case where \( \varphi \) is a linear fractional \( \varphi(z) = \frac{az + b}{cz + d} \) with \( a, b, c, d \in \mathbb{R} \) and \( ad - bc = 1 \). By Proposition 3.15, if \( \vec{v} \) is a vector based at \( z_0 \in \mathbb{H}^2 \),
\[
\|D_{z_0} \varphi(\vec{v})\|_{\text{euc}} = \frac{1}{|cz_0 + d|^2} \|\vec{v}\|_{\text{euc}}.
\]

On the other hand,
\[
\text{Im}(\varphi(z_0)) = \frac{1}{2i} \left( \varphi(z_0) - \overline{\varphi(z_0)} \right) = \frac{1}{2i} \left( \frac{az_0 + b}{cz_0 + d} - \frac{a\bar{z}_0 + b}{c\bar{z}_0 + d} \right)
\]
\[
= \frac{1}{2i} \frac{z_0 - \bar{z}_0}{|cz_0 + d|^2} = \frac{1}{|cz_0 + d|^2} \text{Im}(z_0).
\]

Therefore,
\[
\|D_{z_0} \varphi(\vec{v})\|_{\text{hyp}} = \frac{1}{\text{Im}(\varphi(z_0))} \|D_{z_0} \varphi(\vec{v})\|_{\text{euc}}
\]
\[
= \frac{1}{\text{Im}(z_0)} \|\vec{v}\|_{\text{euc}} = \|\vec{v}\|_{\text{hyp}}.
\]

The argument is essentially identical for an antilinear fractional map \( \varphi(z) = \frac{c\bar{z} + d}{az + b} \). \( \square \)

### 3.6.1. The isotropy property of the hyperbolic plane.

We now show that, like the euclidean plane, the hyperbolic plane \( \mathbb{H}^2 \) is isotropic. Recall that this means that, not only can we send any point \( z_1 \in \mathbb{H}^2 \) to any other point \( z_2 \in \mathbb{H}^2 \) by an isometry \( \varphi \) of \( (\mathbb{H}^2, d_{\text{hyp}}) \), but we can even arrange that \( \varphi \) sends any given direction at \( z_1 \) to any arbitrary direction at \( z_2 \). As a consequence, the hyperbolic plane looks the same at every point and in every possible direction.

**Proposition 3.20.** Let \( \vec{v}_1 \) be a vector based at \( z_1 \in \mathbb{H}^2 \), and let \( \vec{v}_2 \) be a vector based at \( z_2 \in \mathbb{H}^2 \) with \( ||\vec{v}_1||_{\text{hyp}} = ||\vec{v}_2||_{\text{hyp}} \). Then there is an isometry \( \varphi \) of \( (\mathbb{H}^2, d_{\text{hyp}}) \) which sends \( z_1 \) to \( z_2 \) and whose differential map \( D_{z_1} \varphi \) sends \( \vec{v}_1 \) to \( \vec{v}_2 \).

**Proof.** Let \( \theta \in \mathbb{R} \) be the angle from \( \vec{v}_1 \) to \( \vec{v}_2 \) measured in the usual euclidean way, namely after moving \( \vec{v}_1 \) to the point \( z_2 \) by a euclidean translation of \( \mathbb{R}^2 \).

There exists \( c, d \in \mathbb{R} \) such that \( cz_1 + d = e^{-i\frac{\theta}{2}} \). Indeed, finding \( c \) and \( d \) amounts to solving a linear system of two equations. If \( z_1 = x_1 + iy_1 \), one finds \( c = \frac{1}{y_1} \sin \frac{\theta}{2} \) and \( d = \cos \frac{\theta}{2} - \frac{x_1}{y_1} \sin \frac{\theta}{2} \), but the precise value is really irrelevant.

Then, one can find (many) \( a, b \in \mathbb{R} \) such that \( ad - bc = 1 \). This is again a simple linear equation problem, after observing that \( c \) and \( d \) cannot be both equal to 0.

Let \( \varphi_1 \) be the linear fractional defined by \( \varphi_1(z) = \frac{az_1 + b}{cz_1 + d} \). Because of our choice of \( a, b, c, d \), Proposition 3.15 shows that \( D_{z_1} \varphi_1 \) is the complex multiplication by \( e^{i\theta} \), namely the rotation of angle \( \theta \). As a consequence, still comparing angles and directions in the usual euclidean way, \( \vec{v}_3 = D_{z_1} \varphi_1(\vec{v}_1) \) is parallel to \( \vec{v}_2 \) and points in the same direction.

Let \( z_3 = \varphi_1(z_1) \). Let \( \varphi_2 \) be an isometry of \( \mathbb{H}^2 \) sending \( z_3 \) to \( z_2 \). As in our proof of the homogeneity of \( \mathbb{H}^2 \) in Proposition 3.2, we can even arrange that \( \varphi_2 \) is the composition of a homothety with a horizontal translation, so that \( D_{z_3} \varphi_2 \) is a homothety. In particular, \( D_{z_3} \varphi_2 \) sends each vector to one which is parallel to it.

Then, \( \varphi = \varphi_2 \circ \varphi_1 \) sends \( z_1 \) to \( z_2 \), and its differential \( D_{z_1} \varphi = D_{z_3} \varphi_2 \circ D_{z_1} \varphi_1 \) sends \( \vec{v}_1 \) to a vector \( \vec{v}_3' = D_{z_1} \varphi(\vec{v}_1) = D_{z_3} \varphi_2(\vec{v}_3) \) which is based at \( z_2 \) and is parallel to \( \vec{v}_2 \).
By Lemma 3.19,\[
\|\vec{v}_2'\|_{\text{hyp}} = \|D_{z_1} \varphi(\vec{v}_1)\|_{\text{hyp}} = \|\vec{v}_1\|_{\text{hyp}} = \|\vec{v}_2\|_{\text{hyp}}.
\]
The vectors $\vec{v}_2$ and $\vec{v}_2'$ are based at the same point $z_2$, they are parallel, they point in the same direction, and they have the same hyperbolic norm. They consequently must be equal. We therefore have found an isometry $\varphi$ of $(\mathbb{H}^2, d_{\text{hyp}})$ such that $\varphi(z_1) = z_2$ and $D_{z_1} \varphi(\vec{v}_1) = \vec{v}_2$. \hfill \Box

### 3.7. The disk model for the hyperbolic plane

We now describe a new model for the hyperbolic plane, namely another metric space $(X, d)$ which is isometric to $(\mathbb{H}^2, d_{\text{hyp}})$. This model is sometimes more convenient for performing computations. Another side benefit, mathematically less important but not negligible, is that it often leads to prettier pictures, as we will have the opportunity to observe in later chapters.

Let $\mathbb{B}^2$ be the open disk of radius 1 centered at 0 in the complex plane $\mathbb{C}$ namely, in the sense of metric spaces introduced in Section 2.3, the ball $B_{d_{\text{euc}}}((0, 0), 1)$ in the euclidean plane $(\mathbb{R}^2, d_{\text{euc}})$.

For a vector $\vec{v}$ based at the point $z \in \mathbb{B}^2$, define its $\mathbb{B}^2$–norm as
\[
\|\vec{v}\|_{\mathbb{B}^2} = \frac{2}{1 - |z|^2} \|\vec{v}\|_{\text{euc}}
\]
where $\|\vec{v}\|_{\text{euc}}$ is the euclidean norm of $V$. Then, as for the euclidean and hyperbolic plane, define the $\mathbb{B}^2$–length of a piecewise differentiable curve $\gamma$ in $\mathbb{B}^2$ parametrized by $t \mapsto z(t)$, $a \leq t \leq b$, as $\ell_{\mathbb{B}^2}(\gamma) = \int_a^b \|z'(t)\|_{\mathbb{B}^2} dt$. Finally, given two points $P, Q \in \mathbb{B}^2$, define their $\mathbb{B}^2$–distance $d_{\mathbb{B}^2}(P, Q)$ as the infimum of the lengths $\ell_{\mathbb{B}^2}$ as $\gamma$ ranges over all piecewise differentiable curves going from $P$ to $Q$.

Let $\Phi$ be the fractional linear map defined by $\Phi(z) = \frac{-z + i}{z + i}$. Beware that the coefficients of $\Phi$ do not satisfy the usual relation $ad - bc = 1$. This could be achieved by dividing all the coefficients by one of the complex square roots $\pm \sqrt{-2i}$, but the resulting expression would be clumsy and cumbersome.

**Proposition 3.21.** The linear fractional map $\Phi(z) = \frac{-z + i}{z + i}$ induces an isometry from $(\mathbb{H}^2, d_{\text{hyp}})$ to $(\mathbb{B}^2, d_{\mathbb{B}^2})$. 

![Figure 3.7. The disk model for the hyperbolic plane](image)
3.7. THE DISK MODEL FOR THE HYPERBOLIC PLANE

Proof. Note that $|\Phi(z)| = 1$ when $z \in \mathbb{R}$, so that $\Phi$ sends $\mathbb{R} \cup \{\infty\}$ to the unit circle. As a consequence, $\Phi$ sends the upper half-plane $\mathbb{H}^2$ to either the inside or the outside of the unit circle in $\mathbb{C} \cup \{\infty\}$. Since $\Phi(i) = 0$, we conclude that $\Phi(\mathbb{H}^2)$ is equal to the inside $\mathbb{B}^2$ of the unit circle.

Consider the differential $D_z \Phi : \mathbb{C} \to \mathbb{C}$ of $\Phi$ at $z \in \mathbb{H}^2$. By Proposition 3.15 and Complement 3.16, 

$$\|D_z \Phi(v)\|_{\mathbb{B}^2} = \frac{2}{1 - |\Phi(z)|^2} \|D_z \Phi(v)\|_{\text{euc}} = \frac{2}{1 - \frac{|z+i|^2}{|z+i|^2}} - \frac{2i}{(z+i)^2}v$$

$$= \frac{4}{|z+i|^2 - |-z+i|^2} |v| = \frac{4}{(z+i)(\bar{z}-i) - (-z+i)(\bar{z}-i)} |v|$$

$$= \frac{2}{i(\bar{z} - z)} |v| = \frac{1}{\text{Im}(z)} |v| = \|v\|_{\text{hyp}}.$$

From this computation, we conclude that $\Phi$ sends a curve $\gamma$ in $\mathbb{H}^2$ to a curve $\Phi(\gamma)$ in $\mathbb{B}^2$ such that $\ell_{\mathbb{B}^2}(\Phi(\gamma)) = \ell_{\text{hyp}}(\gamma)$.

Taking the infimum of the lengths of such curves, it follows that $d_{\mathbb{B}^2}(\Phi(P), \Phi(Q)) = d_{\text{hyp}}(P, Q)$ for every $P, Q \in \mathbb{H}^2$. In other words, $\Phi$ defines an isometry from $(\mathbb{H}^2, d_{\text{hyp}})$ to $(\mathbb{B}^2, d_{\mathbb{B}^2})$. \[\square\]

In particular, this proves that $d_{\mathbb{B}^2}$ is a metric and not just a semi-metric (namely that $d_{\mathbb{B}^2}(P, Q) = 0$ only when $P = Q$), something which we had implicitly assumed so far.

**Proposition 3.22.** The geodesics of $(\mathbb{B}^2, d_{\mathbb{B}^2})$ are the arcs contained in euclidean circles that are orthogonal to the circle $S^1$ bounding $\mathbb{B}^2$, including straight lines passing through the origin.

**Proof.** Since $\Phi$ is an isometry from $(\mathbb{H}^2, d_{\text{hyp}})$ to $(\mathbb{B}^2, d_{\mathbb{B}^2})$, the geodesics of $(\mathbb{B}^2, d_{\mathbb{B}^2})$ are just the images under $\Phi$ of the geodesics of $(\mathbb{H}^2, d_{\text{hyp}})$.

Because linear fractional sends circles to circles (Proposition 3.18) and respect angles (Corollary 3.17), the result immediately follows from the fact that geodesics of $(\mathbb{H}^2, d_{\text{hyp}})$ are exactly circle arcs in euclidean circles centered on the $x$–axis or, equivalently, orthogonal to this $x$–axis. \[\square\]

**Proposition 3.23.** The isometries of $(\mathbb{B}^2, d_{\mathbb{B}^2})$ are exactly the restrictions to $\mathbb{B}^2$ of all linear and antilinear fractional maps of the form

$$\varphi(z) = \frac{\alpha z + \beta}{\beta z + \alpha} \text{ or } \varphi(z) = \frac{\alpha \bar{z} + \beta}{\beta \bar{z} + \alpha}$$

with $|\alpha|^2 - |\beta|^2 = 1$.

**Proof.** Since $\Phi$ is an isometry from $(\mathbb{H}^2, d_{\text{hyp}})$ to $(\mathbb{B}^2, d_{\mathbb{B}^2})$, the isometries of $(\mathbb{B}^2, d_{\mathbb{B}^2})$ are exactly those maps of the form $\Phi \circ \psi \circ \Phi^{-1}$ where $\psi$ is an isometry of $(\mathbb{H}^2, d_{\text{hyp}})$. \[\square\]
If $\psi$ is a linear fractional map of the form $\psi(z) = \frac{az + b}{cz + d}$ with $a, b, c, d \in \mathbb{R}$ and $ad - bc = 1$, then

$$\Phi \circ \psi \circ \Phi^{-1}(z) = \frac{(ai - b + c + di)z + (-ai - b - c + di)}{(-ai + b + c + di)z + (ai + b - c + di)}$$

is of the form indicated for

$$\alpha = \frac{1}{2}(a + bi - ci + d)$$
$$\beta = \frac{1}{2}(-a + bi + ci + d).$$

Conversely, writing $\alpha + \beta = bi + d$ and $\alpha - \beta = a - ci$ with $a, b, c, d \in \mathbb{R}$, any map $z \mapsto \frac{\alpha z + \beta}{\beta z + \alpha z}$ with $|\alpha|^2 - |\beta|^2 = 1$ is of the form $\Phi \circ \psi \circ \Phi^{-1}$ for some $a, b, c, d \in \mathbb{R}$ with $ad - bc = 1$.

The argument is identical for antilinear fractional maps. \qed

**Exercises for Chapter 3**

**Exercise 3.1.** Rigorously prove that a horizontal translation $\varphi: \mathbb{H}^2 \to \mathbb{H}^2$, defined by the property that $\varphi(x, y) = (x + x_0, y)$ for a given $x_0 \in \mathbb{R}$, is an isometry of the hyperbolic plane $(\mathbb{H}^2, d_{hyp})$.

**Exercise 3.2.** (An explicit formula for the hyperbolic distance). The goal of this exercise is to show that the hyperbolic distance $d_{hyp}(z, z')$ from $z$ to $z' \in \mathbb{H}^2 \subset \mathbb{C}$ is equal to

$$D(z, z') = \log \left| \frac{|z - z'| + |z - z'|}{|z - z'| - |z - z'|} \right|$$

a. Show that $d_{hyp}(z, z') = D(z, z')$ when $z$ and $z'$ are on the same vertical line.

b. Show that $D(\varphi(z), \varphi(z')) = D(z, z')$ for every $z, z' \in \mathbb{H}^2$ when $\varphi: \mathbb{H}^2 \to \mathbb{H}^2$ is a horizontal translation, a homothety or the inversion across the unit circle.

c. Use the proof of Lemma 3.6 to show that $d_{hyp}(z, z') = D(z, z')$ for every $z, z' \in \mathbb{H}^2$.

**Exercise 3.3.** Adapt the proof of Theorem 3.11 to prove that every isometry of the euclidean plane $(\mathbb{R}^2, d_{eucl})$ is of the form $\varphi(z) = z_0 + z e^{i\theta}$ or $\varphi(z) = z_0 + e^{i\theta}z$ for some $z_0 \in \mathbb{C}$ and $\theta \in \mathbb{R}$.

**Exercise 3.4.** (Perpendicular bisector). The **perpendicular bisector** of the two distinct points $P$ and $Q \in \mathbb{H}^2$ is the geodesic $b_{PQ}$ defined as follows. Let $M$ be the midpoint of the geodesic $g$ joining $P$ to $Q$. Then $b_{PQ}$ is the complete geodesic that passes through $M$ and is orthogonal to $g$.

a. Let $\rho$ be the inversion across the euclidean circle that contains $b_{PQ}$. Show that $\rho$ sends the geodesic $g$ to itself, and exchanges $P$ and $Q$.

b. Show that $d_{hyp}(P, R) = d_{hyp}(Q, R)$ for every $R \in b_{PQ}$. Possible hint: Use Part a.

c. Suppose that $P$ and $R$ are on opposite sides of $g_{PQ}$, in the sense that the geodesic $k$ joining $P$ to $R$ meets $b_{PQ}$ in a point $S$. Combine pieces of $k$ and $\rho(k)$ to construct a piecewise differentiable curve $k'$ which goes from $Q$ to $R$, which has the same hyperbolic length as $k$, and which is not geodesic. Conclude that $d_{hyp}(Q, R) < d_{hyp}(P, R)$.

d. As a converse to Part b, show that $d_{hyp}(P, R) \neq d_{hyp}(Q, R)$ whenever $R \notin b_{PQ}$. Possible hint: Use Part c.

**Exercise 3.5.** (Orthogonal projection). Let $g$ be a complete geodesic of $\mathbb{H}^2$, and consider a point $P \in \mathbb{H}^2$.

a. First consider the case where $g$ is a vertical half-line $g = \{(x_0, y) \in \mathbb{R}^2; y > 0\}$. Show that there exists a unique complete geodesic $h$ containing $P$ and orthogonally cutting $g$ at some point $Q$ (namely $h$ and $g$ meet in $Q$ and form an angle of $\frac{\pi}{2}$ there).

b. In the case of a general complete geodesic $g$, show that there exists a unique complete geodesic $h$ containing $P$ and orthogonally cutting $g$ at some point $Q$. Possible hint: Use Lemma 3.6.

c. Show that $Q$ is the point of $g$ that is closest to $P$, in the sense that $d_{hyp}(P, Q') > d_{hyp}(P, Q)$ for every $Q' \in g$ different from $Q$. Possible hint: First consider the case where the geodesic $h$ is equal to a vertical half-line, and where the point $P$ lies above $Q$ on this half-line; then apply Lemma 3.6.
Exercise 3.6 (Hyperbolic rotation around i). For \( \theta \in \mathbb{R} \), consider the fractional linear map defined by

\[
\varphi(z) = \frac{z \cos \frac{\theta}{2} + \sin \frac{\theta}{2}}{-z \sin \frac{\theta}{2} + \cos \frac{\theta}{2}}
\]

a. Show that \( \varphi \) fixes the point \( i \in \mathbb{H}^2 \), and that its differential \( D_i \varphi \) at \( i \) is just the rotation of angle \( \theta \). Hint: Use Lemma 3.15 to compute \( D_i \varphi \).

b. For an arbitrary \( z_0 \in \mathbb{H}^2 \), give a similar formula for the hyperbolic rotation of angle \( \theta \) around the point \( z_0 \), namely for the isometry \( \varphi: \mathbb{H}^2 \to \mathbb{H}^2 \) for which \( \varphi(z_0) = z_0 \) and \( D_{z_0} \varphi \) is the rotation of angle \( \theta \).

Exercise 3.7 (Classification of hyperbolic isometries). Consider an isometry \( \varphi \) of the hyperbolic plane \( (\mathbb{H}^2, d_{hyp}) \), defined by the linear fractional map \( \varphi(z) = \frac{az + b}{cz + d} \) with \( a, b, c, d \in \mathbb{R} \) and \( ad - bc = 1 \).

a. Show that, if \( (a + d)^2 > 4 \), \( \varphi \) has no fixed point in \( \mathbb{H}^2 \) but fixes exactly two points of \( \mathbb{R} \cup \{\infty\} \). Conclude that, in this case, there exists an isometry \( \psi \) of \( (\mathbb{H}^2, d_{hyp}) \) such that \( \psi \circ \varphi \circ \psi^{-1} \) is a homothety \( z \mapsto \lambda z \) with \( \lambda > 0 \). (Hint: Choose \( \psi \) so that it sends the fixed points to 0 and \( \infty \)). Find a relationship between \( \lambda \) and \( (a + d)^2 \). A hyperbolic isometry of this type is said to be \textit{loxodromic}.

b. Show that, if \( (a + d)^2 < 4 \), \( \varphi \) has a unique fixed point in \( \mathbb{H}^2 \). Conclude that, in this case, there is an isometry \( \psi \) of \( (\mathbb{H}^2, d_{hyp}) \) such that \( \psi \circ \varphi \circ \psi^{-1} \) is the linear fractional map of Exercise 3.6 for some \( \theta \in \mathbb{R} \). (Hint: Choose \( \psi \) so that it sends the fixed point to \( i \)). Find a relationship between \( \theta \) and \( (a + d)^2 \). A hyperbolic isometry of this type is said to be \textit{elliptic}.

c. Show that, if \( (a + d)^2 = 4 \) and if \( \varphi \) is not the identity map defined by \( \varphi(z) = z \), then \( \varphi \) has a unique fixed point in \( \mathbb{R} \cup \{\infty\} \). Conclude that, in this case, there is an isometry \( \psi \) of \( (\mathbb{H}^2, d_{hyp}) \) such that \( \psi \circ \varphi \circ \psi^{-1} \) is the horizontal translation \( z \mapsto z + 1 \). (Hint: Choose \( \psi \) so that it sends the fixed point to \( \infty \)). A hyperbolic isometry of this type is said to be \textit{parabolic}.

Exercise 3.8 (Stereographic projection). Let \( S^2 \) be the unit sphere in the 3-dimensional euclidean space \( \mathbb{R}^3 \), consisting of those points \( (x, y, z) \in \mathbb{R}^3 \) such that \( x^2 + y^2 + z^2 = 1 \). Consider the map \( \rho: S^2 \to \mathbb{R}^2 \cup \{\infty\} \) defined as follows: If \( (x, y, z) \neq (0, 0, 1) \), then

\[
\rho(x, y, z) = \left( \frac{x}{1 - z}, \frac{y}{1 - z} \right) \in \mathbb{R}^2;
\]

otherwise, \( \rho(0, 0, 1) = \infty \).

a. Show that, when \( P = (x, y, z) \) is not the ‘North Pole’ \( N = (0, 0, 1) \), its image \( \rho(P) \) is just the point where the line \( NP \) crosses the \( xy \)-plane in \( \mathbb{R}^3 \).

b. Show that \( \rho: S^2 \to \mathbb{R}^2 \cup \{\infty\} \) is continuous at every \( P_0 \in S^2 \). When \( P_0 = N \) so that \( \rho(P_0) = \infty \), this means that, for every large \( \eta > 0 \), there exists a small \( \delta > 0 \) such that \( d_{\text{eucl}}(\rho(P), O) > \eta \) for every \( P \in S^2 \) with \( d_{\text{eucl}}(P, P_0) < \delta \), where \( O \) is the origin in \( \mathbb{R}^2 \). (Compare the calculus definition of infinite limits, as reviewed in Section 1.4 of the tool kit appendix.)

c. Show that the inverse function \( \rho^{-1}: \mathbb{R}^2 \cup \{\infty\} \to S^2 \) is continuous at every \( Q_0 \in \mathbb{R}^2 \cup \{\infty\} \). When \( Q_0 = \infty \) so that \( \rho^{-1}(Q_0) = N \), this means that, for every small \( \epsilon > 0 \), there exists a large \( \eta > 0 \) such that \( d_{\text{eucl}}(\rho^{-1}(Q), N) < \epsilon \) for every \( Q \in \mathbb{R}^2 \) with \( d_{\text{eucl}}(Q, O) > \eta \).

In other words, \( \rho \) is a homeomorphism from \( S^2 \) to \( \mathbb{R}^2 \cup \{\infty\} \). (See Section 6.1 for a definition of homeomorphisms).

Exercise 3.9. Let \( z_0, z_1 \) and \( z_{\infty} \) be three distinct points in the Riemann sphere \( \hat{\mathbb{C}} = \mathbb{C} \cup \{\infty\} \). Show that there exists a unique linear fractional map \( \varphi \) and a unique antilinear fractional map \( \psi \) such that \( \varphi(0) = \psi(0) = z_0 \), \( \varphi(1) = \psi(1) = z_1 \) and \( \varphi(\infty) = \psi(\infty) = z_{\infty} \).

Exercise 3.10.

a. Show that the linear fractional map \( \varphi: \hat{\mathbb{C}} \to \hat{\mathbb{C}} \) defined by \( \varphi(z) = \frac{az + b}{cz + d} \), with \( a, b, c, d \in \mathbb{C} \) such that \( ad - bc = 1 \), is bijective and that its inverse \( \varphi^{-1} \) is the linear fractional map \( \varphi^{-1}(z) = \frac{dz - b}{cz + a} \). Hint: Remember that \( \varphi^{-1}(z) \) is the number \( a \) such that \( \varphi(a) = z \).

b. Give a similar formula for the inverse of the antilinear fractional map \( \psi(z) = \frac{cz + d}{bz + a} \) with \( ad - bc = 1 \).

Exercise 3.11.

a. Show that any linear or antilinear fractional map can be written as the composition of finitely many inversions across circles.

b. Show that, when a linear fractional map is written as the composition of finitely many inversions across circles, the number of inversions is even. (Hint: Corollary 3.17.) Similarly show that, when an antilinear fractional map is written as the composition of finitely many inversions across circles, the number of inversions is odd.

Exercise 3.12 (Linear fractional maps and projective lines). Let the \textit{real projective line} \( \mathbb{RP}^1 \) consist of all 1-dimensional linear subspaces of the vector space \( \mathbb{R}^2 \). Namely, \( \mathbb{RP}^1 \) is the set of all lines \( L \) through the origin in \( \mathbb{R}^2 \). Since such a line \( L \) is determined by its slope \( s \in \mathbb{R} \cup \{\infty\} \), this provides an identification \( \mathbb{RP}^1 \cong \mathbb{R} \cup \{\infty\} \).
3. THE HYPERBOLIC PLANE

a. Let \( \Phi_A: \mathbb{R}^2 \to \mathbb{R}^2 \) be the linear map defined by the matrix \( A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \) with determinant \( ad - bc \) equal to 1.

Similarly, consider the linear fractional map \( \varphi_A: \mathbb{R} \cup \{ \infty \} \to \mathbb{R} \cup \{ \infty \} \) defined by \( \varphi_A(s) = \frac{as + b}{cs + d} \). Show that \( \Phi_A \)

b. Use Part a to show that \( \varphi_{AA'} = \varphi_A \circ \varphi_A' \), where \( AA' \) denotes the product of the matrices \( A \) and \( A' \).

c. Similarly, consider \( \mathbb{C}^2 = \mathbb{C} \times \mathbb{C} \) as a vector space over the field \( \mathbb{C} \), and let the complex projective line \( \mathbb{CP}^1 \)

Exercise 3.13 (Hyperbolic disks). Recall from Section 2.3 that, in a metric space \( (X, d) \), the ball of radius \( r \) centered at \( P \in X \) is \( B_d(P, r) = \{ Q \in X ; d(P, Q) < r \} \).

a. Let \( O \) be the center of the disk model \( \mathbb{H}^2 \) of Section 3.7. Show that the ball \( B_{d_{\mathbb{R}^2}}(O, r) \) in \( \mathbb{R}^2 \) coincides with the euclidean open disk of radius \( \frac{r}{\tan \theta} \) centered at \( O \).

b. Show that every hyperbolic ball \( B_{d_{\mathbb{H}^2}}(P, r) \) in the hyperbolic plane \( \mathbb{H}^2 \) is a euclidean open disk. Possible hint: Use Part a, the isometry \( \Phi: (\mathbb{H}^2, d_{\mathbb{H}^2}) \to (\mathbb{B}^2, d_{\mathbb{B}^2}) \) of Proposition 3.21, Proposition 3.2 and Proposition 3.18.

c. Show that the ball \( B_{d_{\mathbb{H}^2}}(P, r) \) centered at \( P = (x, y) \in \mathbb{H}^2 \) is the euclidean open disk with euclidean radius \( 2y \sinh r \) and with euclidean center \( (x, 2y \cosh r) \). Possible hint: Look at the two points where the boundary of this disk meets the vertical line passing through \( P \).

Exercise 3.14 (Hyperbolic area). If \( D \) is a region in \( \mathbb{H}^2 \), define its hyperbolic area as

\[
\text{Area}_{\mathbb{H}^2}(D) = \int_D \frac{1}{y^2} \, dx \, dy.
\]

a. Let \( \rho: \mathbb{H}^2 \to \mathbb{H}^2 \) be the standard inversion. Show that \( \rho(D) \) has the same hyperbolic area as \( D \). Possible hints: Polar coordinates may be convenient; alternatively, one can use the change of variables formula for double integrals (see Part c).

b. Show that an isometry of \( \mathbb{H}^2 \) sends each region \( D \subset \mathbb{H}^2 \) to a region of the same hyperbolic area.

c. Let \( \Phi \) be the isometry from \( (\mathbb{H}^2, d_{\mathbb{H}^2}) \) to \( (\mathbb{B}^2, d_{\mathbb{B}^2}) \) provided by Proposition 3.21. Show that, for every region \( D \) in \( \mathbb{H}^2 \),

\[
\text{Area}_{\mathbb{H}^2}(D) = \int_{\Phi(D)} \frac{1}{(1 - x^2 - y^2)^2} \, dx \, dy.
\]

It may be convenient to use the change of variables formula for double integrals, which in this case says that

\[
\int_{\Phi(D)} f(u, v) \, du \, dv = \int_D f(\Phi(x, y)) \, |\det D\Phi| \, dx \, dy,
\]

for every function \( f: \Phi(D) \to \mathbb{R} \), and to borrow computations from the proof of Proposition 3.21 to evaluate the determinant \( \det D\Phi \) of the differential map \( D\Phi \).

d. Show that a ball \( B_{d_{\mathbb{H}^2}}(P, r) \) of radius \( r \) in \( \mathbb{H}^2 \) has hyperbolic area

\[
\text{Area}_{\mathbb{H}^2}(B_{d_{\mathbb{H}^2}}(P, r)) = 2\pi (\cosh r - 1) = 4\pi \sinh^2 \frac{r}{2}.
\]

Hint: First consider the case where \( P = (1, 0) = \Phi^{-1}(O) \), and use Part c and the result of Exercise 3.13a.

Exercise 3.15 (Area of hyperbolic triangles). For every \( \theta \) with \( 0 < \theta < \frac{\pi}{2} \), let \( T_\theta \) be the (finite) hyperbolic triangle with vertices \( i, e^{i\theta} = \cos \theta + i \sin \theta \) and \( \infty \). Namely \( T_\theta \) is the region of \( \mathbb{H}^2 \) bounded below by the euclidean circle of radius 1 centered at the origin, on the left by the \( y \)-axis, and on the right by the line \( x = \cos \theta \).

a. Show that the hyperbolic area \( \text{Area}_{\mathbb{H}^2}(T_\theta) \), as defined in Exercise 3.14, is finite.

b. Show that \( \frac{d}{d\theta} \text{Area}_{\mathbb{H}^2}(T_\theta) = -1 \). Conclude that \( \text{Area}_{\mathbb{H}^2}(T_\theta) = \frac{\pi}{2} - \theta \).

c. Let \( T \) be a finite triangle in the hyperbolic plane, namely the region of \( \mathbb{H}^2 \) bounded by the three geodesics joining any two of three distinct points \( P, Q, R \). Let \( \alpha, \beta, \gamma \in [0, \pi] \) be the respective angles of \( T \) at its three vertices. Show that

\[
\text{Area}_{\mathbb{H}^2}(T_\theta) = \pi - \alpha - \beta - \gamma.
\]

Hint: Express this area as a linear combination of the hyperbolic areas of 6 suitably chosen infinite hyperbolic triangles, each isometric to a triangle \( T_\theta \) as in Parts a and b.
EXERCISE 3.16 (Crossratio). The crossratio of four distinct points \(z_1, z_2, z_3, z_4 \in \mathbb{C} = \mathbb{C} \cup \{\infty\}\) is

\[ K(z_1, z_2, z_3, z_4) = \frac{(z_1 - z_3)(z_2 - z_4)}{(z_1 - z_4)(z_2 - z_3)} \in \mathbb{C}, \]

using the straightforward extension by continuity of this formula when one of the \(z_i\) is equal to \(\infty\). For instance, 

\[ K(z_1, z_2, z_3, \infty) = \frac{z_1 - z_3}{z_1 - z_2} \]

Show that \(K(\varphi(z_1), \varphi(z_2), \varphi(z_3), \varphi(z_4))\) is equal to \(K(z_1, z_2, z_3, z_4)\) for every linear fractional map \(\varphi: \mathbb{C} \to \mathbb{C}\), and is equal to the complex conjugate of \(K(z_1, z_2, z_3, z_4)\) for every antilinear fractional map \(\varphi: \mathbb{C} \to \mathbb{C}\). Possible hint: First check the property for a few simple (anti)linear fractional maps, and then apply Lemma 3.12.

EXERCISE 3.17 (Another formula for the hyperbolic distance). Given two distinct points \(z_1, z_2\) of the hyperbolic plane \(\mathbb{H}^2 = \{z \in \mathbb{C}; \text{Im}(z) > 0\}\), let \(x_1, x_2 \in \mathbb{R} = \mathbb{R} \cup \{\infty\}\) be the two end points of the complete geodesic \(g\) passing through \(z_1\) and \(z_2\), in such a way that \(x_1, z_1, z_2, x_2\) occur in this order on \(g\). Show that

\[ d_{\text{hyp}}(z_1, z_2) = \log \frac{(z_1 - x_2)(x_2 - z_1)}{(z_1 - x_1)(x_1 - z_2)}. \]

Possible hint: First consider the case where \(x_1 = 0\) and \(x_2 = \infty\), and then use the invariance property for the crossratio proved in Exercise 3.16.

EXERCISE 3.18. Show that a crossratio formula similar to that to Exercise 3.17 holds in the disk model \((\mathbb{B}^2, d_{\text{hyp}})\) of Section 3.7. Hint: Use the invariance property for the crossratio proved in Exercise 3.16.

EXERCISE 3.19 (The projective model for the hyperbolic plane). Let \(\mathbb{B}^2\) be the open unit disk of Section 3.7. Consider the map \(\Psi: \mathbb{B}^2 \to \mathbb{B}^2\) defined by

\[ \Psi(x, y) = \left(\frac{2x}{1 + x^2 + y^2}, \frac{2y}{1 + x^2 + y^2}\right). \]

a. Show that \(\Psi\) is bijective.

b. Show that, if \(g\) is a complete geodesic of \((\mathbb{B}^2, d_{\text{hyp}})\) which is a circle arc centered on the \(x\)-axis, its image \(\Psi(g)\) is the euclidean line segment with the same end points.

c. Let \(\rho\theta: \mathbb{B}^2 \to \mathbb{B}^2\) be the euclidean rotation of angle \(\theta\) around the origin \(O\). Show that \(\Psi \circ \rho\theta = \rho\theta \circ \Psi\).

d. Combine Parts a and b to show that \(\Psi\) sends each complete geodesic \(g\) of \((\mathbb{B}^2, d_{\text{hyp}})\) to the euclidean line segment with the same end points.

For a vector \(\vec{v}\) based at \(P \in \mathbb{B}^2\), define its projective norm \(\|\vec{v}\|_{\text{proj}} = \|DP\Psi^{-1}(\vec{v})\|_{\mathbb{B}^2}\). For every piecewise differentiable curve \(\gamma\) in \(\mathbb{B}^2\) parametrized by \(t \mapsto \gamma(t), a \leq t \leq b\), define its projective length as \(l_{\text{proj}}(\gamma) = \int_a^b \|\gamma'(t)\|_{\text{proj}} dt\). Finally, consider the new metric \(d_{\text{proj}}\) on \(\mathbb{B}^2\) by the property that \(d_{\text{proj}}(P, Q) = d_{\text{hyp}}(\Psi^{-1}(P), \Psi^{-1}(Q))\) for every \(P, Q \in \mathbb{B}^2\). In particular, \(\Psi\) is now an isometry from \((\mathbb{B}^2, d_{\text{hyp}})\) to \((\mathbb{B}^2, d_{\text{proj}})\).

e. Show that, for every \(P, Q \in \mathbb{B}^2\), the projective distance \(d_{\text{proj}}(P, Q)\) is equal to the infimum of the projective lengths of all piecewise differentiable curves going from \(P\) to \(Q\) in \(\mathbb{B}^2\). Show that this infimum is equal to the projective length of the euclidean line segment from \(P\) to \(Q\).

f. Given a vector \(\vec{v}\) based at \(P \in \mathbb{B}^2\), draw the line \(L\) passing through \(P\) and parallel to \(\vec{v}\), and let \(A, B\) be the two points where it meets the unit circle \(S^1\) bounding \(\mathbb{B}^2\). Show that

\[ \|\vec{v}\|_{\text{proj}} = \frac{d_{\text{euc}}(A, B)}{d_{\text{euc}}(A, P)d_{\text{euc}}(B, P)}\|\vec{v}\|_{\text{euc}}. \]

The computations may be a little easier if one first restricts attention to the case where \(P\) is on the \(x\)-axis, and then use Part c to deduce the general case from this one.

g. For any two distinct \(P, Q \in \mathbb{B}^2\), let \(A, B \in S^1\) be the points where the line \(PQ\) meets the circle \(S^1\), in such a way that \(A, P, Q, B\) occur in this order on the line. Combine Parts e and f to show that

\[ d_{\text{proj}}(P, Q) = \frac{1}{2} \log \frac{d_{\text{euc}}(A, Q)d_{\text{euc}}(P, B)}{d_{\text{euc}}(A, P)d_{\text{euc}}(P, Q)}. \]

(Compare the crossratio formula of Exercise 3.17.)

The metric space \((\mathbb{B}^2, d_{\text{proj}})\), which is isometric to the hyperbolic plane \((\mathbb{H}^2, d_{\text{hyp}})\), is called the projective model or the Cayley–Klein model for the hyperbolic plane. It is closely related to the geometry of the 3-dimensional projective plane \(\mathbb{RP}^2\), defined in close analogy with the projective line \(\mathbb{RP}^1\) of Exercise 3.12 and consisting of all lines passing through the origin in \(\mathbb{R}^3\). The fact that its geodesics are euclidean line segments makes this projective model quite attractive for some problems.
CHAPTER 4

The 2–dimensional sphere

The euclidean plane \((\mathbb{R}^2, d_{\text{euc}})\) and the hyperbolic plane \((\mathbb{H}^2, d_{\text{hyp}})\) have the fundamental property that they are both homogeneous and isotropic. There is another well-known 2–dimensional space which shares this property, namely the 2–dimensional sphere in \(\mathbb{R}^3\).

This is a relatively familiar space but, as will become apparent in later chapters, its geometry is not as fundamental as hyperbolic geometry or, to a lesser extent, euclidean geometry. For this reason, its discussion will be somewhat de-emphasized in this book. We only need a brief description of this space, and of its main properties.

4.1. The 2–dimensional sphere

The 2–dimensional sphere is the set \(S^2 = \{(x, y, z) \in \mathbb{R}^3; x^2 + y^2 + z^2 = 1\}\) consisting of those points in the 3–dimensional space \(\mathbb{R}^3\) which are at euclidean distance 1 from the origin. Namely, \(S^2\) is the euclidean sphere of radius 1 centered at the origin \(O = (0, 0, 0)\).

![Figure 4.1. The 2–dimensional sphere](image)

Given two points \(P, Q \in S^2\), we can consider all piecewise differentiable curves \(\gamma\) that are completely contained in \(S^2\) and join \(P\) to \(Q\). The spherical distance from \(P\) to \(Q\) is defined as the infimum

\[
d_{\text{sph}}(P, Q) = \{\ell_{\text{euc}}(\gamma); \gamma \text{ goes from } P \text{ to } Q \text{ in } S^2\}
\]

of their usual euclidean arc lengths \(\ell_{\text{euc}}(\gamma)\). Here as in Chapter 2, the euclidean arc length of a piecewise differentiable curve \(\gamma\) parametrized by

\[
t \mapsto (x(t), y(t), z(t)), \quad a \leq t \leq b,
\]
is given by
\[ \ell_{\text{euc}}(\gamma) = \int_a^b \sqrt{x'(t)^2 + y'(t)^2 + z'(t)^2} \, dt \]

The definition immediately shows that this spherical distance \( d_{\text{sph}}(P, Q) \) is greater than or equal to the usual euclidean distance \( d_{\text{euc}}(P, Q) \) from \( P \) to \( Q \) in \( \mathbb{R}^3 \). In particular, this proves that \( d_{\text{sph}}(P, Q) = 0 \) only when \( P = Q \). By the same arguments as in Lemma 3.1, \( d_{\text{sph}} \) also satisfies the Symmetry Condition and the Triangle Inequality. This proves that \( d_{\text{sph}} \) is really a metric.

### 4.2. Shortest curves

A **great circle** in the sphere \( S^2 \) is the intersection of \( S^2 \) with a plane passing through the origin. Equivalently, a great circle is a circle of radius 1 contained in \( S^2 \). A **great circle arc** is an arc contained in a great circle.

Elementary geometric considerations show that any two \( P, Q \in S^2 \) can be joined by a great circle arc of length \( \leq \pi \). In addition, this circle arc is unique unless \( P \) and \( Q \) are **antipodal**, namely such that \( Q = (-x, -y, -z) \) if \( P = (x, y, z) \). When \( P \) and \( Q \) are antipodal, there are many great circle arcs of length \( \pi \) going from \( P \) to \( Q \).

**Theorem 4.1.** The geodesics of \((S^2, d_{\text{sph}})\) are exactly the great circle arcs. The shortest curves joining two points \( P, Q \in S^2 \) are exactly the great circle arcs of length \( \leq \pi \) going from \( P \) to \( Q \).

**Proof.** We sketch a proof of this result in Exercise 4.1. \( \square \)

Note that we encounter here a new phenomenon, where a geodesic is not necessarily the shortest curve joining its end points; this happens for every great circle arc of length \( \pi \). Recall that, in the definition of a geodesic \( \gamma \), the part of \( \gamma \) that joins \( P \) to \( Q \) is required to be the shortest curve joining \( P \) to \( Q \) only when \( Q \) is sufficiently close to \( P \).

### 4.3. Isometries

In \( \mathbb{R}^3 \), a rotation \( \varphi \) around a line \( L \) respects euclidean distances and arc lengths. If, in addition, the rotation axis \( L \) passes through the origin \( O \), then \( \varphi \) sends the sphere \( S^2 \) to itself. Consequently, any rotation \( \varphi \) around a line passing through the origin induces an isometry of \((S^2, d_{\text{sph}})\).

These isometries are sufficient to show that the sphere is homogeneous and isotropic, as indicated by the following statement.

**Proposition 4.2.** Given two points \( P, Q \in S^2 \), a vector \( \vec{v} \) tangent to \( S^2 \) at \( P \), and a vector \( \vec{w} \) tangent to \( S^2 \) at \( Q \) such that \( \|\vec{v}\|_{\text{euc}} = \|\vec{w}\|_{\text{euc}} \), there exists a rotation \( \varphi \) around a line passing through the origin \( O \) such that \( \varphi(P) = Q \) and \( D_P\varphi(\vec{v}) = \vec{w} \).

**Proof.** One easily finds a rotation \( \varphi_1 \) such that \( \varphi_1(P) = Q \), for instance one whose axis is orthogonal to the lines \( OP \) and \( OQ \). Then, \( D_P\varphi_1(\vec{v}) \) is a vector tangent to \( S^2 \) at \( \varphi_1(P) = Q \), and whose euclidean length is
\[ \|D_P\varphi_1(\vec{v})\|_{\text{euc}} = \|\vec{v}\|_{\text{euc}} = \|\vec{w}\|_{\text{euc}}. \]

As a consequence, there exists a unique rotation \( \varphi_2 \) around the line \( OQ \) whose differential map \( D_Q\varphi_2 \) sends \( D_P\varphi_1(\vec{v}) \) to \( \vec{w} \). Then the composition \( \varphi = \varphi_2 \circ \varphi_1 \) sends \( P \) to \( Q \), and its differential \( D_P\varphi = D_Q\varphi_2 \circ D_P\varphi_1 \) sends \( \vec{v} \) to \( \vec{w} \).
By a classical property (see Exercise 4.2), the composition \( \varphi = \varphi_2 \circ \varphi_1 \) of two rotations is also a rotation around a line passing through the origin.

We already saw that a rotation \( \varphi \) around a line \( L \) passing through the origin provides an isometry of \( S^2 \). If we compose \( \varphi \) with the reflection across the plane orthogonal to \( L \) at \( O \), we obtain a rotation-reflection.

Note that rotations include the identity map, which is a rotation of angle 0. As a consequence, rotation-reflections also include reflections across a plane passing through the origin.

**Theorem 4.3.** The isometries of \((S^2, d_{\text{sph}})\) are exactly the above rotations and rotation-reflections.

**Proof.** This can be proved by an argument which is very close to the proof of Theorem 3.11. See Exercise 4.4. \( \square \)

**Exercises for Chapter 4**

**Exercise 4.1** (Geodesics of the sphere \( S^2 \)).

(a) Let \( \gamma \) be a piecewise differentiable curve in \( \mathbb{R}^3 \) parametrized by \( t \mapsto \gamma(t), a \leq t \leq b \). For each \( t \), let \( \rho(t), \theta(t) \) and \( \varphi(t) \) be the spherical coordinates of \( \gamma(t) \). Show that the euclidean length of \( \gamma \) is equal to

\[
\ell_{\text{eucl}}(\gamma) = \int_a^b \sqrt{\rho'(t)^2 + \rho(t)^2 \sin^2 \varphi(t) \theta'(t)^2 + \rho(t)^2 \varphi'(t)^2} \, dt.
\]

**Hint:** Remember the formulas expressing rectangular coordinates in terms of spherical coordinates.

(b) In the sphere \( S^2 \), let \( P \) be the point \((0,0,1)\) and let \( Q \) be the point \((x,0,z)\) with \( x \geq 0 \). Let \( \alpha \) be the vertical circle arc going from \( P \) to \( Q \) in \( S^2 \), where the spherical coordinate \( \theta \) is constantly equal to 0. Show that any curve \( \gamma \) going from \( P \) to \( Q \) has euclidean length greater than or equal to that of \( \alpha \).

(c) Show that, if \( P \) and \( Q \) are two points of the sphere \( S^2 \), the shortest curves going from \( P \) to \( Q \) are the great circle arcs of length \( \leq \pi \) going from \( P \) to \( Q \). Possible hint: Use a suitable isometry of \((S^2, d_{\text{sph}})\) to reduce this to the case of Part b.

(d) Show that the geodesics of \( S^2 \) are the great circle arcs.

**Exercise 4.2.** The main goal of this exercise is to show that, in the euclidean space \( \mathbb{R}^3 \), the composition of two rotations whose axes pass through the origin \( O = (0,0,0) \) is also a rotation around a line passing through the origin.

(a) In \( \mathbb{R}^3 \), let \( L \) be a line contained in a plane \( \Pi \). Let \( \Pi' \) be the plane obtained by rotating \( \Pi \) around \( L \) by an angle of \( \frac{\pi}{2} \theta \), and let \( \tau \) and \( \tau' \) be the orthogonal reflections across the planes \( \Pi \) and \( \Pi' \), respectively. Show that the composition \( \tau' \circ \tau \) is the rotation of angle \( \theta \) around \( L \), and that \( \tau \circ \tau' \) is the rotation of angle \( -\theta \) around \( L \). Possible hint: You may find it convenient to consider the restrictions of \( \tau' \circ \tau \) and \( \tau \circ \tau' \) to each plane \( \Pi \) orthogonal to \( L \).

(b) Let \( L \) and \( L' \) be two lines passing through the origin \( O = (0,0,0) \) in \( \mathbb{R}^3 \), and let \( \rho \) and \( \rho' \) be two rotations around \( L \) and \( L' \), respectively. Show that the composition \( \rho \circ \rho' \) is a rotation around a line passing through the origin.

Possible hint: Consider the plane \( \Pi \) containing \( L \) and \( L' \), and use the two properties of Part a.

(c) Show that, in \( \mathbb{R}^3 \), the composition \( \tau_1 \circ \tau_2 \circ \cdots \circ \tau_n \) of an even number of orthogonal reflections \( \tau_i \) across planes passing through \( O \) is a rotation (possibly the identity).

**Exercise 4.3.** The main goal of this exercise is to show that, if \( \tau \) is an orthogonal reflection across a plane \( \Pi \) passing through the origin \( O = (0,0,0) \), and if \( \rho \) is a rotation of angle \( \theta \) around a line \( L \) passing through \( O \), then the composition \( \tau \circ \rho \) is a rotation-reflection. This means that \( \tau \circ \rho \) is also equal to a composition \( \tau' \circ \rho' \) where \( \rho' \) is a rotation across a line \( L' \) passing through \( O \), and where \( \tau' \) is the orthogonal reflection across the plane \( \Pi' \) orthogonal to \( L' \) at \( O \).

Without loss of generality, we can assume that \( L \) is not orthogonal to \( \Pi \), since otherwise we are done. Then, the plane \( \Pi_1 \) orthogonal to \( \Pi \) and containing \( L \) is uniquely determined. Let \( \Pi_2 \) be the image of \( \Pi_1 \) under the rotation of angle \( -\frac{1}{2} \theta \) around \( L \), and let \( \Pi_3 \) be the plane orthogonal to both \( \Pi \) and \( \Pi_1 \) at \( O \). Let \( \tau_1 \) and \( \tau_2 \) be the orthogonal reflections across the planes \( \Pi_1 \) and \( \Pi_2 \), respectively.

(a) Show that \( \tau \circ \tau_1(P) = -P \) for every \( P \in \Pi_3 \) (where \( -P \) denotes the point \((-x,-y,-z)\) when \( P = (x,y,z) \)).

**Drawing a picture might help.**

(b) Show that \( \tau \circ \tau_1 \circ \tau_2(P) = -P \) for every \( P \) in the line \( L' = \Pi_2 \cap \Pi_3 \).
c. Let \( \tau' \) be the orthogonal reflection across the plane \( \Pi' \) orthogonal to \( L' \) at \( O \). Show that \( \rho' = \tau' \circ \tau \circ \tau_1 \circ \tau_2 \) is a rotation around the line \( L' \). Hint: First use the result of Exercise 4.2c to show that \( \rho' \) is a rotation around some line.

d. Show that \( \tau \circ \rho = \tau' \circ \rho' = \rho' \circ \tau' \), so that \( \tau \circ \rho \) is a rotation-reflection. Hint: First use Exercise 4.2a to show that \( \rho = \tau_1 \circ \tau_2 \).

e. Show that, in \( \mathbb{R}^3 \), the composition \( \tau_1 \circ \tau_2 \circ \cdots \circ \tau_{2n+1} \) of an odd number of orthogonal reflections \( \tau_i \) across planes passing through \( O \) is a rotation-reflection. Hint: Use Exercise 4.2c.

Exercise 4.4 (Isometries of the sphere \( (S^2, d_{\text{sph}}) \)).

a. Adapt the proof of Theorem 3.11 to show that every isometry of the sphere \( (S^2, d_{\text{sph}}) \) is a composition of reflections across planes passing through the origin \( O = (0, 0, 0) \).

b. Show that every isometry of the sphere \( (S^2, d_{\text{sph}}) \) is, either a rotation, or a rotation-reflection. Hint: Use the conclusions of Exercises 4.2c and 4.3d.

e. Show that necessarily \( \cos(\alpha + \beta) < -\cos \gamma < \cos(\alpha - \beta) \).

c. Show that the double inequality of Part b is equivalent to the condition that
\[
\max\{\alpha - \beta, \beta - \alpha\} < \pi - \gamma < \min\{\alpha + \beta, 2\pi - (\alpha + \beta)\},
\]
which itself is equivalent to the condition that
\[
\pi < \alpha + \beta + \gamma < \pi + 2\min\{\alpha, \beta, \gamma\}.
\]
Hint: Note that \( -\pi < \alpha - \beta < \pi, 0 < \alpha + \beta < 2\pi \) and \( 0 < \pi - \gamma < \pi \).

d. Combine Parts a, b and c to show that, if
\[
\pi < \alpha + \beta + \gamma < \pi + 2\min\{\alpha, \beta, \gamma\},
\]
there exists a spherical triangle \( T \subset S^2 \) with respective angles \( \alpha, \beta, \gamma \).

e. Show that, if two spherical triangles \( T \) and \( T' \subset S^2 \) have the same angles \( \alpha, \beta, \gamma \), there exists an isometry \( \varphi \) of \( (S^2, d_{\text{sph}}) \) sending \( T \) to \( T' \). Hint: In Part b, there are only two possible unit vectors \( \vec{w} \).

Exercise 4.6 (Area of spherical triangles).

a. In the sphere \( S^2 \), consider two great semi-circles joining the ‘North Pole’ \( (0,0,1) \) to the ‘South Pole’ \( (0,0,-1) \), and making an angle of \( \alpha \) with each other at these poles. Show that the area of the bigon bounded by these two arcs is equal to \( 2\alpha \). Hint: Use spherical coordinates, or a proportionality argument.

b. Let \( T \) a spherical triangle with angles \( \alpha, \beta, \gamma \). Let \( A, B \) and \( C \) be the great circles of \( S^2 \) that contain each of the three edges of \( T \). Show that these great circles subdivide the sphere \( S^2 \) into eight spherical triangles, whose angles are all of the form \( \alpha, \beta, \gamma, \pi - \alpha, \pi - \beta \) or \( \pi - \gamma \).

c. Combine Parts a and b to show that the area of the triangle \( T \) is equal to \( \alpha + \beta + \gamma - \pi \). Hint: Solve a system of linear equations.

d. Show that necessarily \( \pi < \alpha + \beta + \gamma < \pi + 2\min\{\alpha, \beta, \gamma\} \). Hint: First show that \( 0 < \alpha + \beta + \gamma - \pi < 2\alpha \).
CHAPTER 5

Gluing constructions

This chapter and the following one are devoted to the construction of interesting metric spaces which are locally identical to the euclidean plane, the hyperbolic plane or the sphere, but are globally very different. We start with the intuitive idea of gluing together pieces of paper, but then go on with the mathematically rigorous construction of spaces obtained by gluing together the edges of euclidean and hyperbolic polygons. This chapter is concerned with the theoretical aspects of the construction, while the next one will investigate various examples.

5.1. Informal examples: the cylinder and the torus

We first discuss in a very informal way the idea of creating new spaces by gluing. Precise definitions will be rigorously developed in the next section.

If one takes a rectangular piece of paper and glues the top side to the bottom side so as to respect the orientations indicated on Figure 5.1, it is well-known that one gets a cylinder.

![Creating a cylinder from a piece of paper](Figure 5.1)

This paper cylinder can be deformed to many positions in 3-dimensional space but they all have the same metric: As long as we do not stretch the paper, the euclidean arc length of a curve drawn on the cylinder remains constant under deformations, and is actually equal to the arc length of the corresponding pieces of curve in the original rectangle.

One can also try, in addition to gluing the top side to the bottom side, to glue the left side to the right side of the piece of paper. Namely, after gluing the top and bottom sides together to obtain a cylinder, we can glue the left boundary curve of the cylinder to the right one. This is harder to physically realize in 3-space without crumpling the paper but, if we are willing to use rubber instead of paper and to stretch the cylinder in order to put its two sides in contact, one easily sees that one obtains a torus, namely an inner tube or the surface of a donut. See Figure 5.2.

![Creating a cylinder from a piece of paper](Figure 5.1)

Let us try to understand the geometry of this torus, from the point of view of a little bug crawling over it. For instance suppose that, in order to meet her lover, the bug walks from
44 5. GLUING CONSTRUCTIONS

Figure 5.2. Love story on a torus

$P$ to $Q$ along the curve $\gamma$ indicated on the right of Figure 5.2. To measure the distance that it needs to travel, one could consider the euclidean arc length of $\gamma$ for a given position of the torus in 3-dimensional space, but this will depend on the stretching that occurred when moving the torus to that position. However, if we are interested in pre-stretching distances, the natural thing to consider is to decompose the curve $\gamma$ into pieces coming from the original piece of paper, and then take the sum of the lengths of these pieces. For instance, in the situation illustrated on Figure 5.2, the curve $\gamma$ comes from three curves $\gamma_1$, $\gamma_2$ and $\gamma_3$ on the square, in such a way that each $\gamma_i$ goes from a point $P_i$ to a point $Q_i$, and where $P = P_1$, $Q_1$ is glued to $P_2$, $Q_2$ is glued to $P_3$, and $Q_3 = Q$. The distance traveled by our critter friend, as measured on the original piece of paper, is then the sum of the euclidean arc lengths of $\gamma_1$, $\gamma_2$ and $\gamma_3$ on this piece of paper.

In order to introduce some mathematical rigor in this discussion, let us formalize this construction. We begin with the rectangle

$$X = [a, b] \times [c, d] = \{(x, y) \in \mathbb{R}^2; a \leq x \leq b, c \leq y \leq d\}.$$ 

Let $\bar{X}$ be the space obtained from $X$ by doing the gluings indicated. Some points of $\bar{X}$ correspond to exactly one point of $X$ (located in the interior of the rectangle), some points of $\bar{X}$ correspond to two points of $X$ (located on opposite sides of the rectangle), and one point corresponds to four points of $X$ (namely the corners of the rectangle). In other words, each point of $\bar{X}$ is described by a subset of $X$ of one of the following types:

1. the 1-element set $\{(x, y)\}$ with $a < x < b$ and $c < y < d$;
2. the 2-element set $\{(x, c), (x, d)\}$ with $a < x < b$;
3. the 2-element set $\{(a, y), (b, y)\}$ with $c < y < d$;
4. the 4-element set $\{(a, c), (a, d), (b, c), (b, d)\}$.

These subsets form a partition of $X$. This means that every point of $X$ belongs to exactly one such subset.

We could define the distance between points of $\bar{X}$ by taking the infimum of the lengths of curves joining them, as in the example of our little bug walking on the torus. The next section develops a definition which is equivalent to this idea, but is somewhat easier to state and to use.

5.2. Mathematical definition of gluings, and quotient spaces

5.2.1. Partitions. Let $(X, d)$ be a metric space, and consider a partition $\bar{X}$ of $X$. As indicated above, a partition of $X$ is a family $\bar{X}$ of subsets $A \subset X$ such that each point $P \in X$ belongs to one and only one such subset $A$.  

In particular, every element of the set $\tilde{X}$ is a subset $A \subset X$. We can therefore consider that the set $\tilde{X}$ is obtained from $X$ by deciding that, for each subset $A$ of the partition, all the points in the subset $A$ now correspond to a single element of $\tilde{X}$. In other words, all the points of $A$ are now glued together to give a single point in $\tilde{X}$. So the formalism of partitions is a good way to rigorously describe the intuitive idea of gluing points of $X$ together. It takes a while to get used to it, though, since a point of $\tilde{X}$ is also a subset of $X$.

The following notation will often be convenient. If $P \in X$ is a point of $X$, let $\tilde{P} \in \tilde{X}$ denote the corresponding point of $\tilde{X}$ after the gluing. Namely, in the formalism of partitions, $\tilde{P} \subset X$ is the element of the partition $\tilde{X}$ such that $P \in \tilde{P}$.

5.2.2. The quotient semi-metric. We now introduce a distance function on the set $\tilde{X}$, along the lines of the informal discussion of the previous section.

If $P$ and $Q$ are two points of $\tilde{X}$, respectively corresponding to $P$ and $Q \in X$, a discrete walk $w$ from $P$ to $Q$ is a finite sequence $P = P_1, Q_1, P_2, Q_2, P_3, \ldots, Q_{n-1}, P_n, Q_n = Q$ of points of $X$ such that $Q_i = \bar{P}_{i+1}$ for every $i < n$. Namely, such a discrete walk alternates travels in $X$ from $P_i$ to $Q_i$, and jumps from $Q_i$ to a point $P_{i+1}$ that is glued to $Q_i$. The d-length (or just the length if there is no ambiguity on the metric $d$ considered) of a discrete walk $w$ is the sum of the travel distances

$$\ell_d(w) = \sum_{i=1}^{n} d(P_i, Q_i).$$

This is the exact translation of our informal discussion of the little bug walking on the torus, except that we are now requiring the bug to follow a path which is made up of straight line segments in the rectangle. Namely, the bug should be a grasshopper instead of a snail. As it follows a discrete walk, it alternates steps where it hops in $X$ from one point $P_i$ to another point $Q_i$, and steps where it is beamed-up from $Q_i$ to $P_{i+1}$ by the gluing process.

To make the description of a discrete walk $w$ easier to follow, it is convenient to write $P \sim P'$ when $P$ is glued to $P'$, namely when $\bar{P} = \bar{P}'$. (Compare also our discussion of equivalence relations in Exercise 5.2). Then a discrete walk $w$ from $P$ to $Q$ is of the form $P = P_1, Q_1 \sim P_2, Q_2 \sim P_3, \ldots, Q_{n-1} \sim P_n, Q_n = Q$. This makes it a little easier to remember that the consecutive points $Q_i, P_{i+1}$ are glued to each other, while the pairs $P_i, Q_i$ corresponds to a travel of length $d(P_i, Q_i)$ in $X$.

We would like to define a distance function $\bar{d}$ on $\tilde{X}$ by

$$\bar{d}(P, Q) = \inf\{\ell_d(w); w \text{ discrete walk from } P \text{ to } Q\}$$

for any two points $P, Q \in \tilde{X}$.

**Lemma 5.1.** The above number $\bar{d}(P, Q)$ is independent of the choice of the points $P, Q \in X$ used to represent $P, Q \in \tilde{X}$. As a consequence, this defines a function $\bar{d}: \tilde{X} \times \tilde{X} \to \mathbb{R}$.

In addition, this function $\bar{d}$ is a semi-distance on $\tilde{X}$, in the sense that it satisfies the following three conditions.

(1) $\bar{d}(P, Q) \geq 0$ and $\bar{d}(P, \bar{P}) = 0$ for every $\bar{P}, \bar{Q} \in \tilde{X}$ (Non-Negativity);
(2) $\bar{d}(P, Q) = \bar{d}(Q, P)$ for every $P, Q \in \tilde{X}$ (Symmetry);
(3) $\bar{d}(P, \bar{R}) \leq \bar{d}(P, \bar{Q}) + \bar{d}(\bar{Q}, \bar{R})$ for every $P, Q, R \in \tilde{X}$ (Triangle Inequality).

**Proof.** To prove the first statement, consider two other points $P', Q' \in X$ such that $\bar{P}' = \bar{P}$ and $\bar{Q}' = \bar{Q}$. We need to show that $\bar{d}(P', Q') = \bar{d}(P, Q)$.

---

1The original Park City lectures involved pedagogic sound effects to distinguish between these two types of moves.
If $w$ is a discrete walk $P = P_1, Q_1 \sim P_2, Q_2 \sim P_3, \ldots, Q_{n-1} \sim P_n, Q_n = Q$ from $\bar{P}$ to $\bar{Q}$, starting from $P$ and ending at $Q$, we can consider another discrete walk $w'$ of the form $P' = P_0, Q_0 \sim P_1, Q_1 \sim P_2, Q_2 \sim P_3, \ldots, Q_{n-1} \sim P_n, Q_n \sim P_{n+1}, Q_{n+1} = Q'$ by taking $P_0 = Q_0 = P'$ and $P_{n+1} = Q_{n+1} = Q'$. This new discrete walk $w'$ starts at $P'$, ends at $Q'$, and has the same length $\ell(w') = \ell(w)$ as $w$. Taking the infimum over all such discrete walks $w$, we conclude that the “distance” $d(\bar{P}, \bar{Q})$, defined using $P'$ and $Q'$, is less than or equal to $d(\bar{P}, \bar{Q})$ defined using $P$ and $Q$. Exchanging the roles of $P$, $Q$ and $P'$, $Q'$, we similarly obtain that $d(\bar{P}, \bar{Q}) \leq \bar{d}(\bar{P}', \bar{Q}')$, so that $d(\bar{P}, \bar{Q}) = \bar{d}(\bar{P}', \bar{Q}')$.

This proves that the function $\bar{d}: \bar{X} \times \bar{X} \rightarrow \mathbb{R}$ is well-defined.

The Non-Negativity Condition (1) is immediate.

To prove the Symmetry Condition (2), note that every discrete walk $P = P_1, Q_1 \sim P_2, Q_2 \sim P_3, \ldots, Q_{n-1} \sim P_n, Q_n = Q$ going from $\bar{P}$ to $\bar{Q}$ provides a discrete walk $Q = Q_n, P_n \sim Q_{n-1}, P_{n-1} \sim Q_{n-2}, \ldots, P_2 \sim Q_1, P_1 = P$ from $\bar{Q}$ to $\bar{P}$. Since these two discrete walks have the same length, one immediately concludes that $\bar{d}(\bar{P}, \bar{Q}) = \bar{d}(\bar{Q}, \bar{P})$.

Finally, for the Triangle Inequality (3), consider a discrete walk $w$ of the form $P = P_1, Q_1 \sim P_2, Q_2 \sim P_3, \ldots, Q_{n-1} \sim P_n, Q_n = Q$ going from $\bar{P}$ to $\bar{Q}$, and a discrete walk $w'$ of the form $Q = Q_1, R_1 \sim Q_2, \ldots, R_{m-1} \sim Q_m, R_m = R$ going from $\bar{Q}$ to $\bar{R}$. These two discrete walks can be chained together to give a discrete walk $w''$ of the form $P = P_1, Q_1 \sim Q_2, \ldots, Q_{n-1} \sim P_n, Q_n \sim Q_{m}', R_1 \sim Q_2', \ldots, R_{m-1} \sim Q_m', R_m = R$ going from $\bar{P}$ to $\bar{R}$. Since $\ell_d(w'') = \ell_d(w) + \ell(d')$, taking the infimum over all such discrete walks $w$ and $w'$, we conclude that $\bar{d}(\bar{P}, \bar{R}) \leq \bar{d}(\bar{P}, \bar{Q}) + \bar{d}(\bar{Q}, \bar{R})$. □

The only missing property for the semi-distance $\bar{d}$ to be a distance function is that $\bar{d}(\bar{P}, \bar{Q}) = 0$ only when $\bar{P} = \bar{Q}$.

If this property holds, we will say that the partition or gluing process is \textit{proper}. The metric space $(\bar{X}, \bar{d})$ then is the \textit{quotient space} of the metric space $(X, \bar{d})$ by the partition. The distance function $\bar{d}$ is the \textit{quotient metric} induced by $\bar{d}$.

In the case of the torus obtained by gluing the sides of a rectangle, we will see in the next section that the gluing is proper. The same will hold for many examples that we will consider later on. However, there also exist partitions which are not proper. See Exercise 5.1 for an example where each point is glued to at most one other point.

\textbf{5.2.3. The quotient map.} For future reference, the following elementary observation will often be convenient.

Let $\pi: \bar{X} \rightarrow \bar{X}$ be the \textit{quotient map} defined by $\pi(P) = \bar{P}$.

\textbf{Lemma 5.2.} For every $P, Q \in X$,

\[ \bar{d}(\bar{P}, \bar{Q}) \leq d(P, Q). \]

As a consequence, the quotient map $\pi: X \rightarrow \bar{X}$ is continuous.

\textbf{Proof.} The inequality is obtained by consideration of the one-step discrete walk $w$ from $\bar{P}$ to $\bar{Q}$ defined by $P = P_1, Q_1 = Q$. By definition of $\bar{d}$, $\bar{d}(\bar{P}, \bar{Q}) \leq \ell(w) = d(P, Q)$.

The continuity of the quotient map $\pi$ is an immediate consequence of this inequality. □

\textbf{5.3. Gluing the edges of a euclidean polygon}

This section is devoted to the special case where $\bar{X}$ is obtained by gluing together the edges of a polygon $X$, as in the example of the torus obtained by gluing together opposite sides of a rectangle.
5.3.1. Polygons, and edge gluing data. Let $X$ be a *polygon* in the euclidean plane $\mathbb{R}^2$. Namely, $X$ is a region of the euclidean plane whose boundary is decomposed into finitely many line segments, lines and half-lines $E_1, E_2, \ldots, E_n$ meeting only at their end points. We also impose that at most two $E_i$ can meet at any given point.

The line segments, lines and half-lines $E_i$ bounding $X$ are the *edges* of the polygon $X$. The points where two edges meet are its *vertices*.

We require in addition that $X$ and the $E_i$ are *closed*, in the sense that they contain all the points of $\mathbb{R}^2$ that are in their boundary. In this specific case, this is equivalent to the property that the polygon $X$ contains all its edges and all its vertices.

However, we allow $X$ to go to infinity, in the sense that it may be bounded or unbounded in $(\mathbb{R}^2, d_{\text{euc}})$. A subset of a metric space is *bounded* when it is contained in some ball

$$B_{\text{euc}}(P, r) = \{Q \in X; d_{\text{euc}}(P, Q) < r\}$$

with finite radius $r < \infty$. Using the triangle inequality (and changing the radius $r$), one easily sees that this property does not depend on the point $P$ chosen as center of the ball. The subset is *unbounded* when it is not bounded.

Figures 5.3 and 5.4 illustrate a few examples of polygons.

In most cases considered, the polygon $X$ will in addition be *convex* in the sense that, for every $P, Q \in X$, the line segment $[P, Q]$ joining $P$ to $Q$ is contained in $X$. We endow
such a convex polygon with the restriction $d_X$ of the euclidean metric $d_{euc}$, defined by the property that $d_X(P, Q) = d_{euc}(P, Q)$ for every $P, Q \in X$.

Among the polygons $X_1, X_2, \ldots, X_7$ described in Figures 5.3 and 5.4, $X_7$ is the only one that is not convex.

When $X$ is not convex, it is more convenient to consider for $P, Q \in X$ the infimum $d_X(P, Q)$ of the euclidean length $\ell_{euc}(\gamma)$ of all piecewise differentiable curves $\gamma$ joining $P$ and $Q$ and contained in $X$. The fact that the function $d_X$ so defined is a metric on $X$ is immediate, noting that $d_X(P, Q) \geq d_{euc}(P, Q)$. See Exercise 2.10 for an explicit example.

We will call this distance function $d_X$ the euclidean path metric of the polygon $X$. Note that the path metric $d_X$ coincides with the restriction of the euclidean metric $d_{euc}$ when $X$ is convex.

In the general case, the path metric locally coincides with $d_{euc}$ in the sense that every $P \in X$ is the center of a small ball $B_{d_{euc}}(P, \varepsilon)$ such that $d_X(P, Q) = d_{euc}(P, Q)$ for every $Q \in X \cap B_{d_{euc}}(P, \varepsilon)$ (because the line segment $[P, Q]$ is contained in $X$).

We will occasionally allow the polygon $X$ to be made up of several disjoint pieces, so that there exist points $P$ and $Q \in X$ which cannot be joined by a piecewise differentiable curve $\gamma$ completely contained in $X$. In this case, $d_X(P, Q) = +\infty$ by convention. This requires to extend the definition of metrics to allow them to take values in $[0, +\infty]$ but the extension is immediate, provided we use the obvious conventions for inequalities and additions involving infinite numbers ($a \leq \infty$ and $a + \infty = \infty$ for every $a \in [0, \infty]$, etc...).

A polygon for which this does not happen is said to be connected. Namely the polygon $X$ is connected if any two points $P$ and $Q \in X$ can be joined by a piecewise differentiable curve $\gamma$ completely contained in $X$. The reader who is already familiar with some notions of topology will notice that the definition that we are using here is more reminiscent of that of path connectedness; however, the two definitions are equivalent for polygons.

After these preliminaries about polygons, we now describe how to glue the edges of a polygon $X \subset \mathbb{R}^2$ together. For this, we first group these edges into pairs $\{E_1, E_2\}$, $\{E_3, E_4\}$, $\ldots$, $\{E_{2p-1}, E_{2p}\}$ and then, for each such pair $\{E_{2k-1}, E_{2k}\}$, specify an isometry $\varphi_{2k-1} : E_{2k-1} \rightarrow E_{2k}$. Here $\varphi_{2k-1}$ is an isometry for the restriction of the metric $d_X$ on the edges $E_{2k-1}$ and $E_{2k}$, which also coincide with the restrictions of the euclidean metric $d_{euc}$ on these edges. This is equivalent to the property that $\varphi_{2k-1}$ sends each line segment in the edge $E_{2k-1}$ to a line segment of the same length in $E_{2k}$.

Note that, in particular, the edges $E_{2k-1}$ and $E_{2k}$ in a given pair must have the same length, possibly infinite. In general, the isometry $\varphi_{2k-1}$ is then uniquely determined once we know how $\varphi_{2k-1}$ sends which orientation of $E_{2k-1}$ to which orientation of $E_{2k}$. A convenient way to describe this information is to draw arrows on $E_{2k-1}$ and $E_{2k}$ corresponding to these matching orientations. It is also convenient to use a different type of arrow for each pair $\{E_{2k-1}, E_{2k}\}$, so that the pairing can be readily identified on the picture. Figures 5.3 and 5.4 offer some examples.

There is one case where this arrow information is not sufficient, when $E_{2k-1}$ and $E_{2k}$ are both lines (of infinite length), as in the case of the infinite strip $X_5$ of Figure 5.4. In this case, $\varphi_{2k-1} : E_{2k-1} \rightarrow E_{2k}$ is only defined up to translation in one of these lines, and we will need to add extra information to describe $\varphi_{2k-1}$.

The notation will be somewhat simplified if we introduce the isometry $\varphi_{2k} : E_{2k} \rightarrow E_{2k-1}$ defined as the inverse $\varphi_{2k} = \varphi_{2k-1}^{-1}$ of $\varphi_{2k-1} : E_{2k-1} \rightarrow E_{2k}$. In this way, every edge $E_i$ is glued to an edge $E_{i\pm 1}$ by an isometry $\varphi_i : E_i \rightarrow E_{i\pm 1}$, where $\pm 1$ depends on the parity of $i$. 

5.3. GLUING THE EDGES OF A EUCLIDEAN POLYGON

With this data of isometric edge identifications \( \varphi_i: E_i \rightarrow E_{i+1} \), we can now describe the gluing of the edges of the polygon \( X \) by specifying a partition \( \bar{X} \) as follows. Recall that, if \( P \in X \), we denote by \( \bar{P} \in \bar{X} \) the corresponding element of the partition \( \bar{X} \), consisting of all the points of \( X \) that are glued to \( P \). The gluing is then defined as follows:

- if \( P \) is in the interior of the polygon \( X \), then \( \bar{P} \) is glued to no other point and \( \bar{P} = \{P\} \);
- if \( P \) is in a edge \( E_i \) and is not a vertex, then \( \bar{P} \) consists of the two points \( P \in E_i \) and \( \varphi_i(P) \in E_{i+1} \);
- if \( P \) is a vertex, then \( \bar{P} \) consists of \( P \) and of all the vertices of \( X \) of the form \( \varphi_{i_k} \circ \varphi_{i_{k-1}} \circ \cdots \circ \varphi_{i_1}(P) \), where the indices \( i_1, i_2, \ldots, i_k \) are such that \( \varphi_{i_{j-1}} \circ \cdots \circ \varphi_{i_1}(P) \in E_{i_j} \) for every \( j \).

The case of vertices may appear a little complicated at first, but becomes much simpler with practice. Indeed, because each vertex belongs to exactly two edges, there is a simple method to list all elements of \( \bar{X} \) for a vertex \( P \). The key observation is that, when considering a vertex \( \varphi_{i_1} \circ \varphi_{i_{k-1}} \circ \cdots \circ \varphi_{i_1}(P) \) glued to \( P \), we can always assume that the range \( \varphi_{i_1}(E_{i_1}) \) of the gluing map \( \varphi_{i_1} \) is different from the region \( E_{i_{j+1}} \) of \( \varphi_{i_{j+1}} \) (since otherwise \( \varphi_{i_{j+1}} = \varphi_{i_j}^{-1} \), so that these two gluing maps cancel out).

This leads to the following algorithm: Start with \( P_1 = P \), and let \( E_{i_1} \) be one of the two edges containing \( P_1 \). Set \( P_2 = \varphi_{i_1}(P_1) \), and let \( E_{i_2} \) be the edge containing \( P_2 \) that is different from \( \varphi_{i_1}(E_{i_1}) \). Iterating this process provides a sequence of vertices \( P_1, P_2, \ldots, P_j, \ldots \) and edges \( E_{i_1}, E_{i_2}, \ldots, E_{i_j}, \ldots \) such that \( P_j \in E_{i_j}, P_{j+1} = \varphi_{i_j}(P_j) \), and \( E_{i_{j+1}} \) is the edge containing \( P_{j+1} \) that is different from \( \varphi_{i_j}(E_{i_j}) \). Since there are only finitely many vertices, there is an index \( k \) for which \( P_{k+1} = P_j \) for some \( j \leq k \). If \( k \) is the smallest such index, one easily checks that \( P_{k+1} = P_1 \), and that \( \bar{P} = \{P_1, P_2, \ldots, P_k\} \).

For instance, in the example of the rectangle \( X_1 \) of Figure 5.3, the four corners are glued together to form a single point in \( \bar{X}_1 \). For the hexagon \( X_4 \), the vertices of \( X_4 \) project to two points of \( \bar{X}_4 \), each of them corresponding to three vertices of the hexagon. In Figure 5.4, the infinite strip \( X_5 \) has no vertex. The quotient space \( \bar{X}_7 \) has two elements associated to vertices of \( X_7 \), one corresponding to exactly one vertex and another one consisting of two vertices of \( X_7 \).

5.3.2. Edge gluings are proper. Let \( \overline{d}_X \) be the semi-metric on the quotient space \( \bar{X} \) that is defined by the euclidean path metric \( d_X \) of the polygon \( X \subset \mathbb{R}^2 \).

**Theorem 5.3.** If \( \bar{X} \) is obtained from the euclidean polygon \( X \) by gluing together edge pairs by isometries, then the gluing is proper. Namely, the semi-distance \( \overline{d}_X \) induced on \( \bar{X} \) by the metric \( d_X \) of \( X \) is such that \( \overline{d}_X(\bar{P}, \bar{Q}) > 0 \) when \( \bar{P} \neq \bar{Q} \).

The proof of Theorem 5.3 is somewhat long, and is postponed to Section 5.4.

5.3.3. Euclidean surfaces. Let us go back to our informal paper and adhesive tape discussion of the torus \( \bar{X} \), obtained by gluing opposite sides of a rectangle \( X \).

If \( \bar{P} \) is the point of the torus that corresponds to the four corners of the rectangle, it should be intuitively clear (and will be rigorously proved in Lemma 5.5) that a point \( \bar{Q} \in \bar{X} \) is at distance \( < \varepsilon \) of \( \bar{P} \) in \( \bar{X} \) exactly when it corresponds to a point \( Q \in X \) which is at distance \( < \varepsilon \) from one of the corners of the rectangle. As a consequence, for \( \varepsilon \) small enough, the ball \( B_{\overline{d}_X}(\bar{P}, \varepsilon) \) is the image in \( \bar{X} \) of four quarter-disks in \( X \). We know from experience that, if we glue together four paper quarter-disks with (invisible) adhesive tape, we obtain
an object which is undistinguishable from a full disk of the same radius in the euclidean plane.

The same property will hold at a point $P \in \bar{X}$ that is the image of a point $P \in X$ located on a side of the rectangle, not at a corner. Then the ball $B_d(P, \varepsilon)$ is obtained by gluing two half-disks along their diameters, and again has the same metric properties as a disk.

As a consequence, if our little bug crawling on the torus $\bar{X}$ is, in addition, very near-sighted, it will not be able to tell that it is walking on a torus instead of a plane. This may be compared to the (hi)story of other well-known animals who thought for a long time that they were living on a plane, before progressively discovering that they were actually inhabiting a surface with the rough shape of a very large sphere.

When gluing together sectors of paper disks, a crucial property for the result to look like a full disk in the euclidean plane is that the angles of these disk sectors should add up to $2\pi$. As is well-known to anybody who has ever made a birthday hat out of cardboard, the resulting paper construction has a sharp cone point if the sum of the angles is less than $2\pi$. Similarly, it wrinkles if the angles add up to more than $2\pi$. Not unexpectedly, we will encounter the same condition when gluing the edges of a euclidean polygon.

Let us put this informal discussion in a more mathematical framework.

Two metric spaces $(X, d)$ and $(X', d')$ are locally isometric if, for every $P \in X$, there exists an isometry between some ball $B_d(P, \varepsilon)$ centered at $P$ and a ball $B_{d'}(P', \varepsilon)$ in $X'$.

**Theorem 5.4.** Let $(\bar{X}, \bar{d}_X)$ be the quotient metric space obtained from a euclidean polygon $(X, d_X)$ by gluing together pairs of edges of $X$ by isometries. Suppose that the following additional condition holds: For every vertex $P$ of $X$, the angles of $X$ at those vertices $P'$ of $X$ which are glued to $P$ add up to $2\pi$. Then $(\bar{X}, \bar{d}_X)$ is locally isometric to the euclidean plane $(\mathbb{R}^2, d_{euc})$.

Again, although the general idea is exactly the one suggested by the above paper-and-adhesive-tape discussion, the proof of Theorem 5.4 is rather long, with several cases to consider. It is given in the next section.

A metric space $(X, d)$ which is locally isometric to the euclidean plane $(\mathbb{R}^2, d_{euc})$ is a euclidean surface. Equivalently, the metric $d$ is then a euclidean metric.

### 5.4. Proofs of Theorems 5.3 and 5.4

This section is devoted to the proofs of Theorems 5.3 and 5.4. These proofs are not very difficult but a little long, with many cases to consider. They may perhaps be skipped on a first reading, which gives the reader a first opportunity to use the fast-forward command of the remote control.

These are the first really complex proofs that we encounter in this book. It is important to understand these arguments (and many more later) at a level which is higher than a simple manipulation of symbols. With this goal in mind, the reader is strongly encouraged to read them with a piece of paper and pencil in hand, and to draw pictures of the geometric situations involved in order to better follow the explanations.

Throughout this section, $X$ will denote a polygon in the euclidean space $(\mathbb{R}^2, d_{euc})$. We are also given isometries $\varphi_{2k-1}: E_{2k-1} \to E_{2k}$ and $\varphi_{2k} = \varphi_{2k-1}^{-1}: E_{2k} \to E_{2k-1}$ between the edges $E_1, \ldots, E_n$ of $X$. Then $\bar{X}$ is the space obtained from $X$ by performing the corresponding edge gluings, and $\bar{d}_X$ is the semi-metric induced on $\bar{X}$ by the euclidean path metric $d_X$ introduced in the previous section. Recall that, in the more common case where $X$ is convex, $d_X$ is just the restriction of the euclidean metric $d_{euc}$ to $X$. 
We first need to understand the balls

\[ B_{d_X}(\bar{P}, \varepsilon) = \{ \bar{Q} \in \bar{P}; d_X(\bar{P}, \bar{Q}) < \varepsilon \}, \]

at least for \( \varepsilon \) sufficiently small.

### 5.4.1. Small balls in the quotient space \((\bar{X}, \bar{d})\).

**Lemma 5.5.** For every \( \bar{P} \in \bar{X} \), there exists an \( \varepsilon_0 > 0 \) such that, for every \( \varepsilon \leq \varepsilon_0 \) and every \( Q \in X \), the point \( \bar{Q} \in \bar{X} \) is in the ball \( B_{d_X}(\bar{P}, \varepsilon) \) if and only if there is a \( P' \in \bar{P} \) such that \( d_X(P', Q) < \varepsilon \).

We can rephrase this in terms of the quotient map \( \pi : X \to \bar{X} \) sending \( P \in X \) to \( \pi(P) = \bar{P} \in \bar{X} \). Lemma 5.5 states that, for \( \varepsilon \) sufficiently small, the ball \( B_{d_X}(\bar{P}, \varepsilon) \) is exactly the union of the images under \( \pi \) of the balls \( B_{d_X}(P', \varepsilon) \) as \( P' \) ranges over all points of \( \bar{P} \).

For a better understanding of the proof, it may be useful to realize that the ball \( B_{d_X}(\bar{P}, \varepsilon) \) can be significantly larger than the union of the images \( \pi(B_{d_X}(P', \varepsilon)) \) when \( \varepsilon \) is not small. This is illustrated by Figure 5.5 in the case of the torus. In these pictures, \( \bar{P} \) consists of the single point \( P \). The shaded areas represent the balls \( B_{d_X}(\bar{P}, \varepsilon) \) for various values of \( \varepsilon \); in each shaded area, the darker area is the image of the ball \( B_{d_X}(P, \varepsilon) \) under \( \pi \).

![Figure 5.5. Balls \( B_{d}(\bar{P}, \varepsilon) \) in \( \bar{X} \)](image)

**Proof of Lemma 5.5.** Recall from Lemma 5.2 that \( d_X(\bar{P}, \bar{Q}) \leq d_X(P', Q) \) for every \( P' \in \bar{P} \). Therefore, the “if” part of the statement holds without restriction on \( \varepsilon \).

The “only if” part will take more time to prove, as we will need to distinguish cases.

Let \( \bar{P} \) be a point of \( \bar{X} \). For a number \( \varepsilon_0 \) which will be specified later on in function of \( \bar{P} \), we consider a point \( Q \in X \) such that \( d_X(\bar{P}, \bar{Q}) < \varepsilon \leq \varepsilon_0 \). We want to find a point \( P' \in \bar{P} \), namely a point of \( X \) which is glued to \( P \), such that \( d_X(P', Q) < \varepsilon \).

Since \( d_X(\bar{P}, \bar{Q}) < \varepsilon \), there exists a discrete walk \( w \) from \( \bar{P} \) to \( \bar{Q} \), of the form \( P = P_1, Q_1 \sim P_2, Q_2 \sim P_3, \ldots, Q_{n-1} \sim P_n, Q_n = Q \) and whose length is such that \( \ell(w) = \sum_{i=1}^{n} d_X(P_i, Q_i) < \varepsilon \).

We want to prove by induction that, for every \( j \leq n \),

\[
(5.1) \quad \text{there exists } P' \in P \text{ such that } d_X(P', Q_j) \leq \sum_{i=1}^{j} d_X(P_i, Q_i) < \varepsilon.
\]

We can begin the induction with \( j = 1 \), in which case \( (5.1) \) is trivial by taking \( P' = P \).

Suppose as induction hypothesis that \( (5.1) \) holds for \( j \). We want to show that it holds for \( j + 1 \).

For this, we will distinguish cases according to the type of the point \( P \in X \). We will also specify \( \varepsilon_0 \) in each case.
Case 1. \( P \) is in the interior of the polygon \( X \).

We first specify the number \( \varepsilon_0 \) needed in this case. We choose it so that the closed disk of radius \( \varepsilon_0 \) centered at \( P \) is completely contained in the interior of the polygon. Equivalently, every point on the boundary of the polygon is at distance \( > \varepsilon_0 \) from \( P \).

In this case, \( P \) is the only point of \( \bar{P} \).

By the induction hypothesis (5.1) and by choice of \( \varepsilon_0 > \varepsilon \), the point \( Q_j \) is in the interior of the polygon. In particular, it is glued to no other point, so that \( P_{j+1} = Q_j \). Combining the triangle inequality with the induction hypothesis, we conclude that

\[
d_X(P, Q_{j+1}) \leq d_X(P, Q_j) + d_X(P_{j+1}, Q_{j+1}) \leq \sum_{i=1}^{j+1} d_X(P_i, Q_i) < \varepsilon.
\]

This proves (5.1) for \( j + 1 \).

Case 2. \( P \) is on an edge \( E_i \) of the polygon, and not at a vertex.

In this case, \( \bar{P} \) consists of \( P \) and of exactly one other point \( \varphi_i(P) \) on the edge \( E_{i\pm1} \) that is glued to \( E_i \).

Choose \( \varepsilon_1 > 0 \) such that \( P \) is at distance \( > \varepsilon_1 \) from the other edges \( E_j \), with \( j \neq i \), of the polygon. Similarly, let \( \varepsilon_2 \) be such that \( \varphi_i(P) \) is at distance \( > \varepsilon_2 \) from any edge other than the edge \( E_{i\pm1} \) that contains it. Choose \( \varepsilon_0 \) as the smaller of \( \varepsilon_1 \) and \( \varepsilon_2 \).

If \( Q_j = P_{j+1} \), combining the induction hypothesis (5.1) with the triangle inequality gives, as in the case of interior points,

\[
d_X(P', Q_{j+1}) \leq d_X(P', Q_j) + d_X(P_{j+1}, Q_{j+1}) \leq \sum_{i=1}^{j+1} d_X(P_i, Q_i) < \varepsilon,
\]

which proves (5.1) for \( j + 1 \) in this case.

Otherwise, \( Q_j \) and \( P_{j+1} \) are distinct but glued together. Because \( d_X(P', Q_j) < \varepsilon \leq \varepsilon_0 \) and by choice of \( \varepsilon_0 \), these two points cannot be vertices of the polygon, so that one of them is in the edge \( E_i \), and the other one is in the edge \( E_{i\pm1} \) glued to \( E_i \) by the map \( \varphi_i \). In particular, \( P_{j+1} = \varphi_i^{\pm1}(Q_j) \). Set \( P'' = \varphi_i^{\pm1}(P') \). Note that \( P'' \) is just equal to \( P \) or to \( \varphi_i(P) \); in particular it is in \( P \).

We will use the crucial property that the gluing map \( \varphi_i \) respects distances. As a consequence, \( d_X(P'', P_{j+1}) = d_X(P', Q_j) \) so that

\[
d_X(P'', Q_{j+1}) \leq d_X(P'', P_{j+1}) + d_X(P_{j+1}, Q_{j+1})
\leq d_X(P', Q_j) + d_X(P_{j+1}, Q_{j+1}) \leq \sum_{i=1}^{j+1} d_X(P_i, Q_i) < \varepsilon
\]

by induction hypothesis. Again, this proves (5.1) for \( j + 1 \) in this case.

Case 3. \( P \) is a vertex of the polygon.

In this case, \( \bar{P} \) consists of a certain number of vertices \( P' \) of the polygon. Pick \( \varepsilon_0 \) such that every point of \( \bar{P} \) is at distance \( > \varepsilon_0 \) from the edges that do not contain it.

The proof is almost identical to that of Case 2, with only a couple of minor twists.

If \( Q_j = P_{j+1} \), as in the previous two cases, the combination of the induction hypothesis (5.1) and of the triangle inequality shows that (5.1) holds for \( j + 1 \).

If \( Q_j \neq P_{j+1} \), we distinguish cases according to whether \( Q_j \) is a vertex of \( X \) or not. If \( Q_j \) is a vertex, since \( d_X(P', Q_j) < \varepsilon \leq \varepsilon_0 \) by induction hypothesis, this vertex must be \( P' \).
by choice of $\varepsilon_0$. Then, $P'' = P_{j+1}$ is also in $\bar{P}$ since it is glued to $Q_j = P'$, and we therefore found a $P'' \in \bar{P}$ such that

$$d_X(P'', Q_{j+1}) = d_X(P_{j+1}, Q_{j+1}) \leq \sum_{i=1}^{j+1} d_X(P_i, Q_i) < \varepsilon$$

as required.

Otherwise, $Q_j$ is not a vertex, and is glued to $P_{j+1}$ by a gluing map $\varphi_k: E_k \to E_{k+1}$. By choice of $\varepsilon_0$, $P'$ is contained in the edge $E_k$, so that $P'' = \varphi_k(P') \in \bar{P}$ is defined. Since the gluing map $\varphi_k$ respects distances,

$$d_X(P'', Q_{j+1}) \leq d_X(P'', P_{j+1}) + d_X(P_{j+1}, Q_{j+1})$$

$$\leq d_X(P', Q_j) + d_X(P_{j+1}, Q_{j+1}) \leq \sum_{i=1}^{j+1} d_X(P_i, Q_i) < \varepsilon$$

using the fact that the induction hypothesis (5.1) holds for $j$.

Therefore, (5.1) now holds for $j + 1$, as requested.

This completes the proof of (5.1) in all three cases, and for all $j$. The case $j = n$ proves Lemma 5.5 since $Q_n = Q$. $\square$

5.4.2. Proof of Theorem 5.3. We now have the tools needed to prove Theorem 5.3, namely that $d_X(P, Q) > 0$ whenever $\bar{P} \neq \bar{Q}$.

Let $\varepsilon_0$ be associated to $\bar{P}$ by Lemma 5.5.

Because $\bar{X}$ is a partition of $X$, the fact that $\bar{P} \neq \bar{Q}$ implies that these two subsets $\bar{P}$ and $\bar{Q}$ of $X$ are disjoint, namely have no point in common. Since these subsets are finite, it follows that there exists an $\varepsilon_1 > 0$ such that every point of $\bar{P}$ is at distance $> \varepsilon_1$ from every point of $\bar{Q}$. Set $\varepsilon = \min\{\varepsilon_0, \varepsilon_1\} > 0$.

Then $d_X(P, Q) \geq \varepsilon$. Indeed, Lemma 5.5 would otherwise provide a point $P' \in \bar{P}$ such that $d_X(P', Q) < \varepsilon \leq \varepsilon_1$, thereby contradicting the definition of $\varepsilon_1$. $\square$

5.4.3. Proof of Theorem 5.4. For every point $\bar{P} \in \bar{X}$, we need to find an isometry $\psi$ between a small ball $B_{d_X}(\bar{P}, \varepsilon) \subset \bar{X}$ centered at $\bar{P}$ and a small ball $B_{\text{euc}}(P', \varepsilon)$ in the euclidean plane $\mathbb{R}^2$.

Fix an $\varepsilon$ satisfying the conclusions of Lemma 5.5. In addition, choose $\varepsilon$ small enough that each $P' \in \bar{P}$ is at euclidean distance $> 3\varepsilon$ from any edge that does not contain it. In particular, the balls $B_{d_X}(P', \varepsilon)$ are pairwise disjoint, and are disks, half-disks or disk sectors in $\mathbb{R}^2$, according to the type of $P' \in \bar{P}$. Here, a (euclidean) disk sector is one of the two pieces of a euclidean disk $B_{\text{euc}}(P_0, r)$ in $\mathbb{R}^2$ delimited by two half-lines issued from its center $P_0$, as in a slice of pie.

In addition, the triangle inequality shows that the balls $B_{d_X}(P', \varepsilon)$ are at euclidean distance $> \varepsilon$ apart, in the sense that $d_{\text{euc}}(Q', Q'') > \varepsilon$ if $Q' \in B_{d_X}(P', \varepsilon)$ and $Q'' \in B_{d_X}(P'', \varepsilon)$ with $P' \neq P'' \in \bar{P}$.

Lemma 5.5 says that the ball $B_{d_X}(\bar{P}, \varepsilon)$ is obtained by gluing together the balls $B_{d_X}(P', \varepsilon)$ in $X$ centered at the points $P'$ that are glued to $P$. Let

$$B = \bigcup_{P' \in \bar{P}} B_{d_X}(P', \varepsilon)$$

denote the union of these balls.
This subset \( B \subset X \) comes with two natural constructions. The first one is the restriction of the metric \( d_X \). The second one \( d_B \) is similarly defined, but by restricting attention to curves that are contained in \( B \). Namely, \( d_B(Q, Q') \) is the infimum of the euclidean lengths \( \ell_{\text{euc}}(\gamma) \) of all piecewise differentiable curves \( \gamma \) joining \( Q \) to \( Q' \) and contained in \( B \). In particular, \( d_B(Q, Q') = \infty \) if \( Q \) and \( Q' \) are in distinct balls \( B_{d_X}(P', \varepsilon) \) and \( B_{d_X}(P'', \varepsilon) \) of \( B \).

When \( Q \) and \( Q' \) are in the same ball \( B_{d_X}(P', \varepsilon) \) of \( B \), elementary geometry shows that \( d_B(Q, Q') = d_X(Q, Q') \). Indeed, \( B_{d_X}(P', \varepsilon) \) is a disk, a half-disk or a disk sector. Therefore, the only case which requires some thought is that of a disk sector of angle \( \pi \) (since otherwise \( d_B(Q, Q') = d_X(Q, Q') = d_{\text{euc}}(Q, Q') \) by convexity). In this case, one just needs to check that the shortest curve from \( Q \) to \( Q' \) in the polygon \( P \) is, either a single line segment completely contained in \( B_{d_X}(P', \varepsilon) \), or the union of two line segments meeting at the vertex \( P' \) in \( B_{d_X}(P', \varepsilon) \); compare Exercise 2.10.

Since the ball \( B_{d_X}(P, \varepsilon) \) is obtained by performing certain gluings on \( B \), it inherits a quotient semi-metric \( \bar{d}_B \). The advantage of \( \bar{d}_B \) is that it is entirely defined in terms of \( B \), without reference to the rest of \( X \).

**Lemma 5.6.** The metrics \( d_X \) and \( \bar{d}_B \) coincide on the ball \( B_{d_X}(P, \frac{1}{3}\varepsilon) \).

The restriction to the ball of radius \( \frac{1}{3}\varepsilon \) is used to rule out the possibility of a “shortcut” through \( \bar{X} \) making \( \bar{Q} \) and \( \bar{Q}' \) closer in \( \bar{X} \) than in \( B_{\bar{d}_X}(\bar{P}, \varepsilon) = \bar{B} \). The left-hand side of Figure 5.5 provides an example of two such points \( \bar{Q} \) and \( \bar{Q}' \in B_{d_X}(\bar{P}, \varepsilon) \) such that \( \bar{d}_X(\bar{Q}, \bar{Q}') < \bar{d}_B(\bar{Q}, \bar{Q}') \).

**Proof of Lemma 5.6.** By definition of \( d_X \) and \( d_B \), \( d_X(R, R') \leq d_B(R, R') \) for every \( R, R' \in B \). It immediately follows that \( \bar{d}_X(\bar{Q}, \bar{Q}') \leq \bar{d}_B(\bar{Q}, \bar{Q}') \) for every \( \bar{Q}, \bar{Q}' \in B_{d_X}(\bar{P}, \varepsilon) \). Incidentally, this shows that \( \bar{d}_B \) is really a metric, and not just a semi-metric.

To prove the reverse inequality, we need to restrict attention to \( \bar{Q}, \bar{Q}' \in B_{d_X}(\bar{P}, \frac{1}{3}\varepsilon) \). In particular, \( \bar{d}_X(\bar{Q}, \bar{Q}') < \frac{2}{3}\varepsilon \) by the triangle inequality.

Let \( w \) be a discrete walk from \( \bar{Q} \) to \( \bar{Q}' \) in \( X \), of the form \( Q = Q_1, Q_1' \sim Q_2, Q_2' \sim Q_3, \ldots, Q_{n-1}' \sim Q_n, Q_n' = \bar{Q}' \), and whose \( d_X \)-length \( \ell_{d_X}(w) \) is sufficiently close to \( d_X(\bar{Q}, \bar{Q}') \) that \( \ell_{d_X}(w) < \frac{2}{3}\varepsilon \). Then \( \bar{Q}_i = Q_{i+1} \) in \( \bar{X} \) and, using the fact that the quotient map is distance non-increasing (Lemma 5.2),

\[
\sum_{i=1}^{n} \bar{d}_X(\bar{Q}_i, \bar{Q}_{i+1}) \leq \sum_{i=1}^{n} d_X(Q_i, Q_i') < \frac{2}{3}\varepsilon.
\]

A repeated use of the triangle inequality then shows that

\[
\bar{d}_X(\bar{P}, \bar{Q}_i) \leq \bar{d}_X(\bar{P}, \bar{Q}_1) + \sum_{j=1}^{i-1} \bar{d}_X(\bar{Q}_j, \bar{Q}_{j+1}) < \frac{1}{3}\varepsilon + \frac{2}{3}\varepsilon = \varepsilon,
\]

so that all \( \bar{Q}_i \) are in \( B_{d_X}(\bar{P}, \varepsilon) \). Since \( \varepsilon \) satisfies the conclusions of Lemma 5.5, we conclude that all \( Q_i \) and \( Q_i' \) are in the subset \( B \).

If \( P' \neq P'' \in \bar{P} \), then \( d_X(P', P'') \geq 3\varepsilon \) by choice of \( \varepsilon \), and the triangle inequality shows that any point of the ball \( B_{d_X}(P', \varepsilon) \) is at distance \( > \varepsilon \) from any point of \( B_{d_X}(P'', \varepsilon) \). Since \( d_X(Q_i, Q_i') < \frac{1}{3}\varepsilon \), we conclude that \( Q_i \) and \( Q_i' \) are in the same ball \( B_{d_X}(P', \varepsilon) \). In particular, we observed (right above the statement of Lemma 5.6) that \( d_B(Q_i, Q_i') = d_X(Q_i, Q_i') \).

What this shows is that \( w \) is also a discrete walk from \( \bar{Q} \) to \( \bar{Q}' \) in \( B \), whose \( d_B \)-length \( \ell_{d_B}(w) \) is equal to its \( d_X \)-length \( \ell_{d_X}(w) \). As a consequence, \( \bar{d}_B(\bar{Q}, \bar{Q}') \leq \ell_{d_X}(w) \).
Since this holds for every discrete walk \(w\) whose length \(\ell_{d_X}(w)\) is sufficiently close to \(d_X(\bar{Q}, \bar{Q}')\), we conclude that \(d_B(\bar{Q}, \bar{Q}') \leq d_X(\bar{Q}, \bar{Q}')\).

Because we have already shown that the reverse inequality holds, this proves that \(d_B(\bar{Q}, \bar{Q}') = d_X(\bar{Q}, \bar{Q}')\) for every \(\bar{Q}, \bar{Q}' \in B_{d_X}(\bar{P}, \frac{\varepsilon}{3})\).

We are now ready to prove Theorem 5.4. As in the proof of Lemma 5.5, we will distinguish cases according to the type of the point \(P \in X\) corresponding to \(P \in X\).

**Case 1. \(P\) is in the interior of the polygon \(X\).**

In particular, \(P\) is glued to no other point, so that \(\bar{P}\) consists only of \(P\). Then \(B = B_{d_X}(P, \varepsilon)\) and, by our choice of \(\varepsilon\), the ball \(B_{d_X}(P, \varepsilon)\) is completely contained in the interior of \(X\). In particular, the ball \(B_{d_X}(P, \varepsilon) \subset X\) is the same as the euclidean ball \(B_{d_{euc}}(P, \varepsilon) \subset \mathbb{R}^2\), an open disk in the euclidean plane \(\mathbb{R}^2\). Also, there are no gluings between distinct points of \(B = B_{d_X}(P, \varepsilon)\), so that every \(\bar{Q} \in B_{d_X}(\bar{P}, \varepsilon)\) corresponds to exactly one point \(Q \in B_{d_X}(P, \varepsilon)\).

Define \(\psi : B_{d_X}(P, \varepsilon) \to B_{d_{euc}}(P, \varepsilon)\) by the property that \(\psi(\bar{Q}) = Q\) for every \(\bar{Q} \in B_{d_X}(\bar{P}, \varepsilon)\).

The map \(\psi\) may not be an isometry over the whole ball, but we claim that

\[d_{euc}(\psi(\bar{Q}), \psi(\bar{Q}')) = d_X(\bar{Q}, \bar{Q}')\]

for every \(\bar{Q}, \bar{Q}' \in B_{d_X}(\bar{P}, \frac{\varepsilon}{3})\). Indeed, \(d_X(\bar{Q}, \bar{Q}') = d_B(\bar{Q}, \bar{Q}')\) by Lemma 5.6. Since there are no gluings in \(B\), one easily sees that \(d_B(\bar{Q}, \bar{Q}') = d_B(\bar{Q}, \bar{Q}')\) (see Exercise 5.3). Finally, \(d_B(Q, Q') = d_{euc}(\psi(\bar{Q}), \psi(\bar{Q}'))\) by convexity of the ball \(B = B_{d_X}(P, \varepsilon)\).

This proves that the restriction of \(\psi\) to the ball \(B_{d_X}(\bar{P}, \frac{\varepsilon}{3})\) is an isometry from \((B_{d_X}(\bar{P}, \frac{\varepsilon}{3}), d_X)\) to the euclidean disk \((B_{d_{euc}}(P, \frac{\varepsilon}{3}), d_{euc})\), as requested.

Having completed the analysis in Case 1, we now directly jump to the most complex case.

**Case 2. \(P\) is a vertex of the polygon \(X\).**

Write \(\bar{P} = \{P_1, P_2, \ldots, P_k\}\) with \(P = P_1\). Namely, \(P_1, P_2, \ldots, P_k\) are the vertices of \(X\) that are glued to \(P\). Lemma 5.5 says that the ball \(B_{d_X}(\bar{P}, \varepsilon)\) in \(X\) is the image under the quotient map \(\pi : X \to \bar{X}\) of the union \(B\) of the balls \(B_{d_X}(P_1, \varepsilon), B_{d_X}(P_2, \varepsilon), \ldots, B_{d_X}(P_k, \varepsilon)\) in \(X\).

Because of our choice of \(\varepsilon\), each of the balls \(B_{d_X}(P_j, \varepsilon)\) in the metric space \((X, d_X)\) is a disk sector of radius \(\varepsilon\) in \(\mathbb{R}^2\), and these disk sectors are pairwise disjoint. We now need to work harder than in the previous case to rearrange these disk sectors into a full disk.

![The polygon X](image)

**Figure 5.6.** Gluing vertices together

Each \(P_j\) belongs to exactly two edges \(E_{ij}\) and \(E'_{ij}\). As in our description of vertex gluings at the end of Section 5.3.1, we can choose the indexings so that, for every \(j\) with \(1 \leq j \leq k\),
the gluing map $\varphi_{ij}$ sends the vertex $P_j$ to $P_{j+1}$ and the edge $E_{ij}$ to $E_{ij+1}$, with the convention that $P_{k+1} = P_1$ and $j'_{k+1} = j'_1$.

We will construct our isometry $\psi: B_{d_X}(\bar{P}, \varepsilon) \to B_{d_{euc}}(P', \varepsilon)$ piecewise from suitable isometries $\psi_j$ of $(\mathbb{R}^2, d_{euc})$. For this, we use the following elementary property, which we list as a lemma for future reference.

**Lemma 5.7.** Let $\varphi: g \to g'$ be an isometry between two line segments, half-lines or full lines $g$ and $g'$ in $\mathbb{R}^2$. Then, $\varphi$ extends to an isometry $\varphi: \mathbb{R}^2 \to \mathbb{R}^2$ of $(\mathbb{R}^2, d_{euc})$.

In addition, if we choose one side of $g$ and another side for $g'$, we can arrange that $\varphi$ sends the selected side of $g$ to the one selected for $g'$. The isometry $\varphi$ is then uniquely determined by these properties.

Lemma 5.7 is an immediate consequence of the classification of isometries of $(\mathbb{R}^2, d_{euc})$ provided by Proposition 2.3.

In particular, for every $j$, we can extend the gluing map $\varphi_{ij}: E_{ij} \to E_{ij+1}$ to an isometry $\varphi_{ij}: \mathbb{R}^2 \to \mathbb{R}^2$ of $(\mathbb{R}^2, d_{euc})$ that sends the polygon $X$ to the side of $E_{ij+1}$ that is opposite $X$.

To define the $\psi_j$, we begin with any isometry $\psi_1$ of $(\mathbb{R}^2, d_{euc})$, and inductively define

$$\psi_{j+1} = \psi_j \circ \varphi_{j+1}^{-1} = \varphi_1 \circ \varphi_{i_1}^{-1} \circ \varphi_{i_2}^{-1} \circ \cdots \circ \varphi_{i_j}^{-1}.$$

By induction on $j$ and because $P_{j+1} = \varphi_j(P_j)$, the map $\psi_j$ sends the vertex $P_j$ to the same point $P' = \psi_1(P)$ for every $j$. In particular, the isometry $\psi_j$ sends the disk sector $B_{d_X}(P_j, \varepsilon)$ to a disk sector of the disk $B_{d_{euc}}(P', \varepsilon)$. Similarly, the image of the edge $E_{ij+1} = \varphi_{ij+1}(E_{ij})$ under $\psi_{j+1}$ is equal to the image of $E_{ij}$ under $\psi_j$. By definition of the extension of $\psi_{ij}$ to an isometry of $\mathbb{R}^2$, the two disk sectors $\psi_j(B_{d_X}(P_j, \varepsilon))$ and $\psi_{j+1}(B_{d_X}(P_{j+1}, \varepsilon))$ sit on opposite sides of $\psi_j(E_{ij}) = \psi_{j+1}(E_{ij+1})$. It follows that the disk sectors $\psi_j(B_{d_X}(P_j, \varepsilon))$ all fit side-by-side and in order of increasing $j$ around their common vertex $P'$. See Figure 5.6.

It is now time to use the hypothesis that the internal angles of the polygon $X$ at the vertices $P_1, P_2, \ldots, P_k \in \bar{P}$ add up to $2\pi$. This implies that the disk sector $\psi_{k+1}(B_{d_X}(P_{k+1}, \varepsilon)) = \psi_{k-1}(B_{d_X}(P_1, \varepsilon))$ is equal to $\psi_1(B_{d_X}(P_1, \varepsilon))$. In particular, the two isometries $\psi_{k+1}$ and $\psi_1$ of $(\mathbb{R}^2, d_{euc})$ send $P_1 = P_{k+1}$ to the same point $P'$, send the edge $E_{ik+1} = E_{i_1}$ to the same line segment or half-line issued from $P'$, and send a side of $E_{ik+1} = E_{i_1}$ to the same side of $\psi_{k+1}(E_{ik+1}) = \psi_1(E_{i_1})$. By the uniqueness part of Lemma 5.7, it follows that $\psi_{k+1} = \psi_1$.

Finally note that, when $Q \in E_{ij}$ is glued to $Q' = \varphi_{ij}(Q) \in E_{ij+1}$, then $\psi_j(Q) = \psi_{j+1}(Q')$. We can therefore define a map

$$\psi: B_{d_X}(\bar{P}, \varepsilon) \to B_{d_{euc}}(P', \varepsilon)$$

by the property that $\psi(Q)$ is equal to $\psi_j(Q)$ whenever $Q \in B_{d}(P_j, \varepsilon)$. The above considerations show that $\psi$ is well-defined.

We will show that $\psi$ induces an isometry between the corresponding balls of radius $\frac{1}{2} \varepsilon$.

For this, consider two points $Q, Q' \in B_{d_X}(\bar{P}, \frac{1}{2}\varepsilon)$. By Lemma 5.6 and by the Triangle Inequality, $d_B(Q, Q') = d_X(Q, Q') < \frac{3}{2} \varepsilon$. Let $w$ be a discrete walk from $Q$ to $Q'$ in $B$, of the form $Q = Q_1, Q'_1 \sim Q_2, Q'_2 \sim Q_3, \ldots, Q'_n \sim Q_n, Q' = Q'$, and whose $d_B$-length $\ell_{d_B}(w)$ is sufficiently close to $d_B(Q, Q')$ that $\ell_{d_B}(w) < \frac{3}{2} \varepsilon$. In particular, each $d_B(Q_i, Q'_i)$ is finite, so that $Q_i$ and $Q'_i$ belong to the same ball $B_{d_X}(P_j, \varepsilon)$. As a consequence,

$$d_{euc}(\psi(Q_i), \psi(Q'_i)) = d_{euc}(\psi_j(Q_i), \psi_j(Q'_i))$$

$$= d_{euc}(Q_i, Q'_i) \leq d_B(Q_i, Q'_i)$$
since each $\psi_{j_i}$ is a euclidean isometry. Then, by iterating the Triangle Inequality and using the fact that $Q'_i = Q_{i+1}$,

$$d_{\text{euc}}(\psi(Q), \psi(Q)) \leq \sum_{i=1}^{n-1} d_{\text{euc}}(\psi(Q_i), \psi(Q'_i)) \leq \sum_{i=1}^{n-1} d_{B}(Q_i, Q'_i) = \ell_{d_B}(w).$$

Since this holds for every discrete walk $w$ from $Q$ to $Q'$ in $B$ whose length is sufficiently close to $d_B(Q, Q')$, we conclude that

$$\tag{5.2} d_{\text{euc}}(\psi(Q), \psi(Q)) \leq d_B(Q, Q').$$

Conversely, let $\gamma$ be the oriented line segment from $\psi(Q)$ to $\psi(Q')$ in the disk $B_{d_{\text{euc}}}(P', \frac{1}{3}\varepsilon)$. Recall that $B_{d_{\text{euc}}}(P', \frac{1}{3}\varepsilon)$ is decomposed into the disk sectors $\psi_j(B_d(P_j, \frac{1}{6}\varepsilon))$. We can therefore split $\gamma$ into line segments $\gamma_1, \gamma_2, \ldots, \gamma_n$, in this order, such that each $\gamma_i$ is contained in a disk sector $\psi_{j_i}(B_d(P_{j_i}, \frac{1}{6}\varepsilon))$.

In the disk sector $B_d(P_{j_i}, \frac{1}{6}\varepsilon) \subset X$, consider the oriented line segment $\gamma_i' = \psi_{j_i}^{-1}(\gamma_i)$ corresponding to $\gamma_i$. If the end points of $\gamma_i'$ are labelled so that $\gamma_i'$ goes from $Q_i$ to $Q'_i$, we now have a discrete walk $w$ from $Q$ to $Q'$ of the form $Q = Q_1, Q'_1 \sim Q_2, Q'_2 \sim Q_3, \ldots, Q'_{n-1} \sim Q_n, Q'_n = Q'$, of $d_B$-length

$$\ell_{d_B}(w) = \sum_{i=1}^{n} d_{B}(Q_i, Q'_i) = \sum_{i=1}^{n} \ell_{\text{euc}}(\gamma_i') = \sum_{i=1}^{n} \ell_{\text{euc}}(\gamma_i) = \ell_{\text{euc}}(\gamma) = d_{\text{euc}}(\psi(Q), \psi(Q)).$$

It follows that

$$\tag{5.3} d_B(Q, Q') \leq d_{\text{euc}}(\psi(Q), \psi(Q)).$$

Combining the inequalities (5.2) and (5.3), we conclude that

$$d_X(Q, Q') = d_B(Q, Q') = d_{\text{euc}}(\psi(Q), \psi(Q))$$

for every $Q, Q' \in B_{d_X}(P, \frac{1}{3}\varepsilon)$. In other words, $\psi$ induces an isometry from the ball $(B_{d_X}(P, \frac{1}{3}\varepsilon), d_X)$ to the ball $(B_{d_{\text{euc}}}(P', \frac{1}{3}\varepsilon), d_{\text{euc}})$.

This concludes our discussion of Case 2, where $P$ is a vertex of $X$. We have one case left to consider.

**Case 3.** $P$ is in an edge of the polygon $X$, but is not a vertex.

The proof is identical to that of Case 2. Actually, it can even be considered as a special case of Case 2, by viewing $P$ and the point $P'$ that is glued to it as vertices of $X$ where the internal angle is equal to $\pi$.

This concludes the proof of Theorem 5.4. \qed

### 5.5. Gluing hyperbolic and spherical polygons

Before applying Theorems 5.3 and 5.4 to specific examples, let us look at the key ingredients of their proof in more details. For Theorem 5.3 (and Lemma 5.5 before), in addition to standard properties of metric spaces, we mostly used the fact that the maps $\varphi_i$ gluing one edge of the polygon to another respected distances between points in this edges. A critical component of the proof of Theorem 5.4 was Lemma 5.7.
5.5.1. Hyperbolic polygons. All these properties have straightforward analogues in the hyperbolic plane \((\mathbb{H}^2, d_{\text{hyp}})\), provided we use the appropriate translation. For instance, the euclidean metric \(d_{\text{euc}}\) just needs to be replaced by the hyperbolic metric \(d_{\text{hyp}}\), euclidean isometries by hyperbolic isometries, line segments and lines by geodesics, etc. Consequently, our results automatically extend to the hyperbolic context.

The only point which requires some thought is the following property, which replaces Lemma 5.7.

**Lemma 5.8.** In the hyperbolic plane \((\mathbb{H}^2, d_{\text{hyp}})\), let \(\varphi: g \rightarrow g'\) be an isometry between two geodesics \(g\) and \(g'\). Then, \(\varphi\) extends to an isometry \(\varphi: \mathbb{H}^2 \rightarrow \mathbb{H}^2\) of \((\mathbb{H}^2, d_{\text{hyp}})\).

In addition, if we choose one side of \(g\) and another side for \(g'\), we can arrange that \(\varphi\) sends the selected side of \(g\) to the one selected for \(g'\). The isometry \(\varphi\) is then uniquely determined by these properties.

**Proof.** Pick a point \(P \in g\) and a non-zero vector \(\vec{v}\) tangent to \(g\) at \(P\). In particular, \(\vec{v}\) defines an orientation for \(g\), which we can transport through \(\varphi\) to obtain an orientation of \(g'\). At the point \(P' = \varphi(P) \in g\), let \(\vec{v}'\) be the vector tangent to \(g'\) in the direction of this orientation, and such that \(\|\vec{v}'\|_{\text{hyp}} = \|\vec{v}\|_{\text{hyp}}\). Proposition 3.20, which shows that \((\mathbb{H}^2, d_{\text{hyp}})\) is isotropic, provides a hyperbolic isometry \(\psi: \mathbb{H}^2 \rightarrow \mathbb{H}^2\) such that \(\psi(P) = P'\) and \(D_P\psi(\vec{v}) = \vec{v}'\). In particular, \(\psi\) sends \(g\) to the geodesic which is tangent to \(D_P\psi(\vec{v}) = \vec{v}'\) at \(\psi(P) = P'\), namely \(g'\).

The restriction of \(\psi\) to the geodesic \(g\) is an isometry \(g \rightarrow g'\) which sends \(P\) to the same point \(P'\) as the isometry \(\varphi\), and sends the orientation of \(g\) to the same orientation as \(\varphi\). It follows that \(\psi\) coincides with \(\varphi\) on \(g\). In other words, the isometry \(\psi: \mathbb{H}^2 \rightarrow \mathbb{H}^2\) extends \(\varphi: g \rightarrow g'\).

If \(\psi\) sends the selected side of \(g\) to the selected side of \(g'\), we are done. Otherwise, let \(\rho\) be the hyperbolic reflection across \(g'\), namely the isometry of \((\mathbb{H}^2, d_{\text{hyp}})\) induced by the inversion across the euclidean circle containing \(g'\). Because \(\rho\) fixes every point of \(g'\) and exchanges its two sides, the hyperbolic isometry \(\varphi = \rho \circ \psi\) now has the required properties.

The uniqueness easily follows from Lemma 3.10. \qed

![Figure 5.7. A few hyperbolic polygons](image)

We are now ready to carry out our automatic translation from euclidean to hyperbolic geometry.

Let \(X\) be a **polygon** in the hyperbolic plane \((\mathbb{H}^2, d_{\text{hyp}})\). Namely, \(X\) is a region in \(\mathbb{H}^2\) whose boundary in \(\mathbb{H}^2\) is decomposed into finitely many hyperbolic geodesics \(E_1, E_2, \ldots, E_n\) meeting only at their end points. When we consider \(X\) as a subset of \(\mathbb{R}^2\), its boundary in \(\mathbb{R}^2\) may also include finitely many intervals in the real line \(\mathbb{R}\) bounding the hyperbolic plane.
\[ \mathbb{H}^2 \text{ in } \mathbb{R}^2 = \mathbb{C}; \text{ if this is the case, note that } X \text{ will be unbounded for the hyperbolic metric } d_{\text{hyp}}. \]

We require in addition that \( X \) and the \( E_i \) contain all those points of \( \mathbb{H}^2 \) that are in their boundary. Namely, \( X \) and the \( E_i \) are \textit{closed} in \( \mathbb{H}^2 \), although not necessarily in \( \mathbb{R}^2 \).

The geodesics \( E_i \) bounding \( X \) are the \textit{edges} of the polygon \( X \). The points where two edges meet are its \textit{vertices}. As in the euclidean case, we require that only two edges meet at any given vertex.

Figure 5.7 offers a few examples. In this figure, \( X_1 \) is a hyperbolic octagon, with eight edges and eight vertices; it is bounded for the hyperbolic metric \( d_{\text{hyp}} \). The hyperbolic polygon \( X_2 \) is an infinite strip, with two edges and no vertex; it touches the line \( \mathbb{R} \) along two disjoint intervals, and is unbounded for the hyperbolic metric \( d_{\text{hyp}} \) (although it is bounded for the euclidean metric of \( \mathbb{R}^2 \)). The hyperbolic quadrilateral \( X_3 \) is delimited by our edges, has not vertex in \( \mathbb{H}^2 \), and touches \( \mathbb{R} = \mathbb{R} \cup \{ \infty \} \) along four points, one of which is \( \infty \). We will meet these three hyperbolic polygons again in Sections 6.2, 6.4.2 and 6.5, respectively.

For such a hyperbolic polygon \( X \), we can then introduce edge gluing data by, first, grouping the edges together in pairs \( \{E_1, E_2\}, \{E_3, E_4\}, \ldots, \{E_{2p-1}, E_{2p}\} \) and then, for each such pair \( \{E_{2k-1}, E_{2k}\} \), by specifying an isometry \( \varphi_{2k-1} : E_{2k-1} \rightarrow E_{2k} \). Here \( \varphi_{2k-1} \) is required to be an isometry for the hyperbolic distance \( d_{\text{hyp}} \).

As in the euclidean case, the edges \( E_{2k-1} \) and \( E_{2k} \) in the same pair must have the same hyperbolic length, possibly infinite. In general, the isometry \( \varphi_{2k-1} \) is then uniquely determined once we know how \( \varphi_{2k-1} \) sends an orientation of \( E_{2k-1} \) to an orientation of \( E_{2k} \), and we will often describe this information by drawing matching arrows on \( E_{2k-1} \) and \( E_{2k} \).

The case where drawing arrows is not sufficient to specify \( \varphi_{2k-1} \) sends an orientation of \( E_{2k-1} \) to an orientation of \( E_{2k} \), and we will often describe this information by drawing matching arrows on \( E_{2k-1} \) and \( E_{2k} \). The case where drawing arrows is not sufficient to specify \( \varphi_{2k-1} \) when \( E_{2k-1} \) and \( E_{2k} \) are complete geodesics of \( \mathbb{H}^2 \), namely full euclidean semi-circles centered on the real line.

As in the euclidean case, we endow \( X \) with the \textit{path metric} \( d_X \) for which \( d_X (P, Q) \) is the infimum of the hyperbolic lengths of all piecewise differentiable curves joining \( P \) to \( Q \) in \( X \). When \( X \) is \textit{convex}, in the sense that the geodesic arc joining any two \( P, Q \in X \) is contained in \( X \), the metric \( d_X \) clearly coincides with the restriction of the hyperbolic metric \( d_{\text{hyp}} \).

**Theorem 5.9.** If \( X \) is obtained from the hyperbolic polygon \( X \) by gluing pairs of its edges by isometries, then the gluing is proper. Namely, the semi-distance \( d_X \) induced on \( X \) by the path metric \( d_X \) of \( X \) is such that \( d_X (P, Q) > 0 \) when \( P \neq Q \).

**Proof.** The proof is identical to that of Theorem 5.3. Just follow each step of that proof, using the appropriate translation. \( \square \)

**Theorem 5.10.** Let \( (\bar{X}, \bar{d}_X) \) be the quotient metric space obtained from a hyperbolic polygon \( (X, d_X) \) by gluing together pairs of edges of \( X \) by hyperbolic isometries. Suppose that the following additional condition holds: For every vertex \( P \) of \( X \), the angles of \( X \) at those vertices \( P' \) of \( X \) which are glued to \( P \) add up to \( 2\pi \). Then \( (\bar{X}, \bar{d}) \) is locally isometric to the hyperbolic plane \( (\mathbb{H}^2, d_{\text{hyp}}) \).

**Proof.** The proof is identical to that of Theorem 5.4, provided that we replace Lemma 5.7 by Lemma 5.8. \( \square \)

A metric space \( (X, d) \) which is locally isometric to the hyperbolic plane \( (\mathbb{H}^2, d_{\text{euc}}) \) is a \textit{hyperbolic surface}. Equivalently, the metric \( d \) is then a \textit{hyperbolic metric}. 

5.5.2. Spherical polygons. The same properties also generalize to polygons in the sphere \((S^2, d_{sph})\).

A **polygon** in the sphere \((S^2, d_{sph})\) is a region \(X\) of \(S^2\) whose boundary is decomposed into finitely many geodesics \(E_1, E_2, \ldots, E_n\) meeting only at their end points. These \(E_i\) are the **edges** of the polygon \(X\), and the points where they meet are its **vertices**. As before, we require that \(X\) contains all its edges and vertices, and that every edge contains its end points. Also, exactly two edges meet at a given vertex.

We endow \(X\) with the **path metric** \(d_X\) for which \(d_X(P, Q)\) is the infimum of the euclidean lengths of all piecewise differentiable curves joining \(P\) to \(Q\) in \(X \subset S^2 \subset \mathbb{R}^3\). When \(X\) is **convex**, in the sense that any two \(P, Q \in X\) can be joined by a geodesic arc of \(S^2\) of length \(< \pi\) which is completely contained in \(X\), the metric \(d_X\) clearly coincides with the restriction of the spherical metric \(d_{sph}\).

After grouping the edges together in pairs \(\{E_1, E_2\}, \{E_3, E_4\}, \ldots, \{E_{2p-1}, E_{2p}\}\), the gluing data used consists of isometries \(\varphi_{2k-1}: E_{2k-1} \to E_{2k}\).

As in the hyperbolic case, the key property is the following extension of Lemma 5.7.

**Lemma 5.11.** In the sphere \((S^2, d_{sph})\), let \(\varphi: g \to g'\) be an isometry between two geodesics \(g\) and \(g'\). Then, \(\varphi\) extends to an isometry \(\varphi: S^2 \to S^2\) of \((S^2, d_{sph})\).

In addition, if we choose one side of \(g\) and another side for \(g'\), we can arrange that \(\varphi\) sends the selected side of \(g\) to the one selected for \(g'\). The isometry \(\varphi\) is then uniquely determined by these properties.

**Proof.** Since geodesics of \((S^2, d_{sph})\) are great circle arcs (Theorem 4.1), this is easily proved by elementary arguments in 3–dimensional euclidean geometry. □

As before, we endow the spherical polygon \(X\) with the path metric \(d_X\) for which \(d_X(P, Q)\) is the infimum of the spherical lengths of all piecewise differentiable curves joining \(P\) to \(Q\) in \(X\).

Then, by replacing Lemma 5.7 by Lemma 5.11, the proofs of Theorems 5.3 and 5.4 immediately extend to the spherical context, and give the following two results.

**Theorem 5.12.** If \(\bar{X}\) is obtained from the spherical polygon \(X\) by gluing pairs of its edges by isometries, then the gluing is proper. Namely, the semi-distance \(d_X\) induced on \(\bar{X}\) by the path metric \(d_X\) is really a metric. □

**Theorem 5.13.** Let \((\bar{X}, d_{\bar{X}})\) be the quotient metric space obtained from a spherical polygon \((X, d_X)\) by gluing together pairs of edges of \(X\) by hyperbolic isometries. Suppose that the following additional condition holds: For every vertex \(P\) of \(X\), the angles of \(X\) at those vertices \(P'\) of \(X\) which are glued to \(P\) add up to \(2\pi\). Then \((\bar{X}, d_{\bar{X}})\) is locally isometric to the sphere \((S^2, d_{sph})\). □

**Exercises for Chapter 5**

**Exercise 5.1.** Let \(X\) be the closed interval \([0, 1]\) in \(\mathbb{R}\). Let \(\bar{X}\) be the partition consisting of all the subsets \(\{m/n, m/n\}\) where \(m, n \in \mathbb{N}\) are integers such that \(m\) is odd and \(1 < m < 2^n\), and of all one-element subsets \(\{P\}\) where \(P \in [0, 1]\) is not of the form \(P = \frac{m}{2^n}\) or \(\frac{m}{2^n}\) as above. Let \(d_{\text{euc}}\) be the quotient semi-metric induced on \(\bar{X}\) by the usual metric \(d_{\text{euc}}(P, Q) = |P - Q|\) of \(\bar{X} = [0, 1]\).

a. Show that, for every \(P \in [0, 1]\) and every \(\varepsilon > 0\), there exists \(Q_1, Q_2 \in [0, 1]\) such that \(\bar{Q}_1 = \bar{Q}_2\) in \(\bar{X}\), \(d_{\text{euc}}(0, Q_1) < \varepsilon\), and \(d_{\text{euc}}(P, Q_2) < \varepsilon\).

b. Show that \(d_{\text{euc}}(0, P) = 0\) for every \(P \in \bar{X}\). In particular, the semi-metric \(d_{\text{euc}}\) is not a metric, and the gluing is not proper.

c. Show that \(d_{\text{euc}}(P, Q) = 0\) for every \(P, Q \in \bar{X}\).
EXERCISE 5.2 (Equivalence relations and partitions). A relation on a set $X$ is just a subset $\mathcal{R}$ of the product $X \times Y$. One way to think of this is that $\mathcal{R}$ describes a certain property involving two points of $X$; namely, $P, Q \in X$ satisfy this property exactly when the pair $(P, Q)$ is an element of $\mathcal{R}$. To emphasize this interpretation, we write $P \sim Q$ to say that $(P, Q) \in \mathcal{R}$.

An equivalence relation is a relation such that

(i) $P \sim P$ for every $P \in X$ ( Reflexivity Property);
(ii) if $P \sim Q$, then $Q \sim P$ (Symmetry Property);
(iii) if $P \sim Q$ and $Q \sim R$, then $P \sim Q$ (Transitivity Property).

a. Given an equivalence relation, define the equivalence class of $P \in X$ as

$$\bar{P} = \{Q \in X; P \sim Q\}.$$ 

Show that, as $P$ ranges over all points of $X$, the family of the equivalence classes $\bar{P}$ is a partition of $X$.

b. Conversely, let $\bar{P}$ be a partition of the set $X$ and, as usual, let $P \in X$ denote the subset that contains $P \in X$. Define a relation on $X$ by the property that $P \sim Q$ exactly when $P$ and $Q$ belong to the same subset $\bar{P} = Q$ of the partition. Show that $\sim$ is an equivalence relation.

EXERCISE 5.3 (Trivial gluing). Let $(X, d)$ be a metric space, and consider the trivial gluing where the partition $\bar{X}$ consists only of the one-element subsets $\bar{P} = \{P\}$. In other words, no two distinct elements of $X$ are glued together in $\bar{X}$. Let $(\bar{X}, \bar{d})$ be the resulting quotient semi-metric space. Rigorously prove, using the definition of the quotient semi-metric $\bar{d}$ in terms of discrete walks, that the quotient map $\pi: X \to \bar{X}$ defined by $\pi(P) = \bar{P}$ is an isometry from $(X, d)$ to $(\bar{X}, \bar{d})$.

EXERCISE 5.4 (Iterated gluings). Let $\bar{X}$ be a partition of the metric space $(X, d)$, and let $\tilde{X}$ be a partition of the quotient space $X/\sim$. Let $\tilde{d}$ be the semi-metric induced by $d$ on $X$, and let $\bar{d}$ be the semi-metric induced by $\tilde{d}$ on $\tilde{X}$.

a. If $P \in X$, the element $\bar{P} = \bar{\tilde{P}} \in \tilde{X}$ is a family of subsets of $X$, and we can consider their union $\bar{P} \subset X$. Show that the subsets $\bar{P}$ form a partition $X$ of $X$.

b. Let $\varphi: \bar{X} \to \tilde{X}$ be the map defined by the property that $\varphi(\bar{P}) = \tilde{P}$ for every $P \in X$. Show that $\varphi$ is bijective.

c. Let $w$ be a discrete walk $P = P_1, Q_1 \sim P_2, \ldots, Q_{n-1} \sim P_n, Q_n = Q$ from $P$ to $Q$ in $\bar{X}$. Show that $\tilde{P} = \bar{P}_1, \bar{Q}_1 \sim \bar{P}_2, \ldots, \bar{Q}_{n-1} \sim \bar{P}_n, \bar{Q}_n = \bar{Q}$ from $\tilde{P}$ to $\tilde{Q}$ in $\tilde{X}$ whose length is such that $\ell_{\tilde{d}}(w) \leq \ell_{\bar{d}}(w)$. (Beware that same symbol $\sim$ is used to refer to gluing with respect to the partition $\bar{X}$ in the first case, and with respect to $\tilde{X}$ in the second instance.) Conclude that, if $d$ is the quotient semi-metric induced by $d$ on $X$, then $\bar{d}(\tilde{P}, \tilde{Q}) \leq \tilde{d}(\bar{P}, \bar{Q})$ for every $P, Q \in X$.

d. Given a small $\varepsilon > 0$, let $\bar{w}$ be a discrete walk $\bar{P} = \bar{P}_1, \bar{Q}_1 \sim \bar{P}_2, \ldots, \bar{Q}_{n-1} \sim \bar{P}_n, \bar{Q}_n = \bar{Q}$ from $\tilde{P}$ to $\tilde{Q}$ in $\tilde{X}$ whose length is sufficiently close to $\tilde{d}(\bar{P}, \bar{Q})$ that $\ell_{\tilde{d}}(w) \leq \tilde{d}(\bar{P}, \bar{Q}) + \varepsilon$. Similarly, for every $i$, choose a discrete walk $\bar{w}_i$ from $\bar{P}_i$ to $\bar{Q}_i$, consisting of $P_i = P_{i-1}, Q_{i-1} \sim P_i, Q_i = Q$ whose length is sufficiently close to $\tilde{d}(\bar{P}_i, \bar{Q}_i)$ that $\ell_{\tilde{d}}(w_i) \leq \tilde{d}(\bar{P}_i, \bar{Q}_i) + \frac{\varepsilon}{2^i}$. Show that the $w_i$ can be chained together to form a discrete walk $\bar{w}$ from $\bar{P}$ to $\bar{Q}$ in $\tilde{X}$ such that $\ell_{\tilde{d}}(w) \leq \ell_{\tilde{d}}(w_i) + \frac{\varepsilon}{2^i}$. Conclude that $\bar{d}(\tilde{P}, \tilde{Q}) \leq \tilde{d}(\bar{P}, \bar{Q}) + \varepsilon$.

e. Show that $\varphi$ is an isometry from $(X, \tilde{d})$ to $(\bar{X}, \bar{d})$.

In other words, a two-step gluing construction yields the same quotient semi-metric space as gluing everything together in one single action.

EXERCISE 5.5. Let $X$ be the interval $[0, 2\pi] \subset \mathbb{R}$, and let $\tilde{X}$ be the partition consisting of the two-element subset $\{0, 2\pi\}$ and of all the one-element subsets $\{P\}$ with $P \in (0, 2\pi)$. Let $\bar{d}$ be the quotient semi-metric induced on $\bar{X}$ by the usual metric $d(P, Q) = |Q - P|$ of $X = [0, 2\pi]$. In other words, $(\bar{X}, \bar{d})$ is obtained by gluing together the two end points of the interval $X = (0, 2\pi)$ endowed with the metric $d$.

Let $S^1 = \{(x, y) \in \mathbb{R}^2; x^2 + y^2 = 1\}$ be the unit circle in the euclidean plane, endowed with the metric $d_{S^1}$ for which $d_{S^1}(P, Q)$ is the infimum of the euclidean arc lengths $\ell_{\text{arc}}(\gamma)$ of all piecewise differentiable curves $\gamma$ going from $P$ to $Q$ in $S^1$. Consider the map $\varphi: X = [0, 2\pi] \to S^1$ defined by $\varphi(t) = (\cos t, \sin t)$.

a. Show that, if $\pi: \bar{X} \to X$ denotes the quotient map, there exists a unique map $\bar{\varphi}: \bar{X} \to S^1$ such that $\varphi = \bar{\varphi} \circ \pi$.

b. Show that $\bar{\varphi}$ is bijective.

c. Show that, for every discrete walk $w$ from $\bar{P}$ to $\bar{Q}$ in $\bar{X}$, there exists a piecewise differentiable curve $\gamma$ going from $\varphi(\bar{P})$ to $\varphi(\bar{Q})$ in $S^1$ whose length $\ell_{\text{arc}}(\gamma)$ is equal to the length $\ell_{\tilde{d}}(w)$ of $w$. Conclude that $d_{S^1}(\bar{\varphi}(\bar{P}), \bar{\varphi}(\bar{Q})) \leq \bar{d}(\bar{P}, \bar{Q})$ for every $\bar{P}, \bar{Q} \in \bar{X}$.

d. Combine these results to show that $\varphi$ is an isometry from $(\bar{X}, \bar{d})$ to $(S^1, d_{S^1})$. 
Exercise 5.6. In the euclidean plane $\mathbb{R}^2$, let $D_1$, $D_2$ and $D_3$ be three disjoint euclidean disk sectors of radius $r$ and respective angles $\theta_1$, $\theta_2$ and $\theta_3$ with $\theta_1 + \theta_2 + \theta_3 \leq \pi$. Let $\bar{X}$ be the quotient space obtained from $X = D_1 \cup D_2$ by isometrically gluing one edge of $D_1$ to an edge of $D_2$, sending the vertex of $D_1$ to the vertex of $D_2$. Show that, if $d_X$ and $d_{D_3}$ are the euclidean path metrics defined as in Section 5.3.1, the quotient space $(\bar{X}, d_X)$ is isometric to $(\bar{D}_3, d_{D_3})$. Hint: Copy parts of the proof of Theorem 5.4.

Exercise 5.7 (Euclidean cones). Let $D_1$, $D_2$, $\ldots$, $D_n$ be $n$ disjoint euclidean disk sectors with radius $r$ and with respective angles $\theta_1$, $\theta_2$, $\ldots$, $\theta_n$, respectively, and let $E_i$ and $E_i'$ denote the two edges of $D_i$. Isometrically glue each edge $E_i$ to $E_{i+1}'$, sending the vertex of $D_i$ to the vertex of $D_i + 1$ and counting indices modulo $n$ (so that $E_{n+1}' = E_1'$). Show that the resulting quotient space $(\bar{X}, d_X)$ depends only on the radius $r$ and on the angle sum $\theta = \sum_{i=1}^{n} \theta_i$. Namely, if $D_1'$, $D_2'$, $\ldots$, $D_n'$ is another family of $n'$ disjoint euclidean disk sectors of the same radius $r$ and with respective angles $\theta_1'$, $\theta_2'$, $\ldots$, $\theta_{n'}'$, with $\sum_{i=1}^{n'} \theta_i' = \sum_{i=1}^{n} \theta_i$, and if these disk sectors are glued together as above, then the resulting quotient space $(\bar{X}', d_{X'})$ is isometric to $(\bar{X}, d_X)$. Hint: Use the results of Exercises 5.4 and 5.6 to reduce the problem to the case where $n = n'$ and $\theta_i = \theta_i'$ for every $i$, and to make sure that the order of the gluings does not matter.

The space $(\bar{X}, d_X)$ of Exercise 5.7 is a **euclidean cone** with radius $r$ and cone angle $\theta$. Figure 5.8 represents a few examples. Note the different shape according to whether the cone angle $\theta$ is less than, equal to, or more than $2\pi$. When $\theta$ is equal to $2\pi$, the cone is of course isometric to a euclidean disk.

![Figure 5.8. Three euclidean cones](image)

Exercise 5.8 (Surfaces with cone singularities). Let $(\bar{X}, d_X)$ be the quotient metric space obtained from an euclidean polygon $(X, d_X)$ by isometrically gluing together its edges. Show that, for every $P \in \bar{X}$, there exists a radius $r$ such that the ball $B_{\bar{d}_X}(P, r)$ is isometric to a euclidean cone, defined as in Exercise 5.7.

A space $(\bar{X}, d_X)$ satisfying the conclusions of Exercise 5.8 is a **euclidean surface with cone singularities**. One can similarly define hyperbolic and spherical surfaces with cone singularities.

Exercise 5.9.

a. Let $(C, d)$ be a euclidean cone with center $P_0$, radius $r$ and cone angle $\theta$ as in Exercise 5.7. Show that, for every $r' < r$, the ‘circle’

$$S_d(P_0, r') = \{ P \in C ; d(P, P_0) = r' \}$$

is a closed curve, whose length $\ell_d(S_d(P_0, r'))$ in the sense of Exercise 2.11 is equal to $\theta r'$.

b. Conclude that the angle condition of Theorem 5.4 is necessary for its conclusion to hold. Namely, if, when isometrically gluing together the sides of a euclidean polygon $X$, there is a vertex $P$ such that the angles of $X$ at those vertices $P'$ which are glued to $P$ do not add up to $2\pi$, then the quotient space $(\bar{X}, d)$ is not locally isometric to the euclidean plane $(\mathbb{R}^2, d_{eucl})$.
CHAPTER 6

Gluing examples

After suffering through the long proofs of Section 5.4, we can now harvest the fruit of our labor, and apply to a few examples the technology that we have built in Chapter 5.

6.1. Some euclidean surfaces

We begin by revisiting, in a more rigorous setting, the example of the torus that we had informally discussed in Section 5.1.

6.1.1. Euclidean tori from rectangles and parallelograms. Let \( \mathbf{X}_1 \) be the rectangle \([a, b] \times [c, d]\), consisting of those \((x, y) \in \mathbb{R}^2\) such that \(a \leq x \leq b\) and \(c \leq y \leq d\). Glue the bottom edge \( E_1 = [a, b] \times \{c\} \) to the top edge \( E_2 = [a, b] \times \{d\} \) by the isometry \( \varphi_1 : [a, b] \times \{c\} \rightarrow [a, b] \times \{d\} \) defined by \( \varphi_1(x, c) = (x, d) \), and glue the left edge \( E_3 = \{a\} \times [c, d] \) to the right edge \( E_4 = \{b\} \times [c, d] \) by \( \varphi_3 : \{a\} \times [c, d] \rightarrow \{b\} \times [c, d] \) defined by \( \varphi_3(a, y) = (b, y) \). Namely, we consider the edge gluing that already appeared in Section 5.1, and which we reproduce in Figure 6.1.

![Figure 6.1. Gluing opposite sides of a rectangle](image)

With these edge identifications, the four vertices of the rectangle are glued together to form a single point \( \bar{P} \) of the quotient space \( \overline{\mathbf{X}_1} \). The sum of the angles of \( \mathbf{X}_1 \) at these vertices is

\[
\frac{\pi}{2} + \frac{\pi}{2} + \frac{\pi}{2} + \frac{\pi}{2} = 2\pi.
\]

We can therefore apply Theorems 5.3 and 5.4. Note that \( \mathbf{X}_1 \) is convex, so that the path metric \( d_{\mathbf{X}_1} \) coincides with the restricton of the euclidean metric \( d_{\text{euc}} \) of \( \mathbb{R}^2 \). Then Theorems 5.3 and 5.4 show that the metric space \((\overline{\mathbf{X}_1}, \bar{d}_{\mathbf{X}_1})\) is locally isometric to the euclidean metric of the euclidean plane \((\mathbb{R}^2, d_{\text{euc}})\).

This is our first rigorous example of a euclidean surface.

In our informal discussion of this example in Section 5.1, we explained how \( \overline{\mathbf{X}_1} \) can be identified to the torus illustrated on the right of Figure 6.1 if we are willing to stretch the metric. The mathematically rigorous way to express this property is use the language of topology, and to say that the space \( \overline{\mathbf{X}_1} \) is homeomorphic to the torus.

A **homeomorphism** from a metric space \((\mathbf{X}, d)\) to another metric space \((\mathbf{X}', d')\) is a bijection \( \varphi : \mathbf{X} \rightarrow \mathbf{X}' \) such that both \( \varphi \) and its inverse \( \varphi^{-1} \) are continuous. (See Section 1.1
in the Tool Kit at the end of this book for the definition of bijections and inverse maps.) The homeomorphism \( \varphi \) can be used as a dictionary between \( X \) and \( X' \) to translate back and forth every property involving limits and continuity. For instance, a sequence \( P_1, P_2, \ldots, P_n, \ldots \) converges to the point \( P_\infty \) in \( X \) if and only if its image \( \varphi(P_1), \varphi(P_2), \ldots, \varphi(P_n), \ldots \) converges to \( \varphi(P_\infty) \) in \( X' \). As a consequence, if the metric spaces \((X, d)\) are \((X', d')\) are **homeomorphic**, in the sense that there exists a homeomorphism between them, then \((X, d)\) and \((X', d')\) share exactly the same limit and continuity properties.

An example of homeomorphism is provided by an isometry from \((X, d)\) to \((X', d')\). However, a general homeomorphism \( \varphi: X \to X' \) is much more general, in the sense that the distance \( d'(\varphi(P), \varphi(Q)) \) may be very different from \( d(P, Q) \). The only requirement is that \( d'(\varphi(P), \varphi(Q)) \) is small exactly when \( d(P, Q) \) is small (in a sense quantified with the appropriate \( \varepsilon \) and \( \delta \)).

In general, we will keep our discussion of homeomorphisms at a very informal level. However, we should perhaps go through at least one example in detail.

Let the 2–dimensional torus be the surface \( \mathbb{T}^2 \) of the 3–dimensional space \( \mathbb{R}^3 \) obtained by revolving about the \( z \)-axis the circle in the \( xz \)-plane that is centered at the point \((R, 0, 0)\) and has radius \( r < R \). We consider \( \mathbb{T}^2 \) as a metric space by endowing it with the restriction of the 3–dimensional metric \( d_{\text{eucl}} \) of \( \mathbb{R}^3 \). Different choices of \( r \) and \( R \) give different subsets of \( \mathbb{R}^3 \), but these are easily seen to be homeomorphic.

**Lemma 6.1.** Let \((X_1, \tilde{d}_{X_1})\) be the quotient metric space obtained from the rectangle \( X_1 = [a, b] \times [c, d] \) by gluing together opposite edges by euclidean translations. Then \( X_1 \) is homeomorphic to the 2–dimensional torus \( \mathbb{T}^2 \).

**Proof.** To simplify the notation, we restrict attention to the case of the square \( X_1 = [-\pi, \pi] \times [-\pi, \pi] \). However, the argument straightforwardly extends to general rectangles \([a, b] \times [c, d]\) by suitable rescaling of the variables.

Let \( \rho: X_1 \to \mathbb{T}^2 \) be the map defined by

\[
\rho(\theta, \varphi) = ((R + r \cos \varphi) \cos \theta, (R + r \cos \varphi) \sin \theta, r \sin \varphi).
\]

Geometrically, \( \rho(\theta, \varphi) \) is obtained by rotating by an angle of \( \theta \) around the \( z \)-axis the point \((R + r \cos \varphi, 0, r \sin \varphi)\) of the circle in the \( xz \)-plane with center \((R, 0, 0)\) and radius \( r \). From this description it is immediate that \( \rho(\theta, \varphi) = \rho(\theta', \varphi') \) exactly when \((\theta, \varphi)\) and \((\theta', \varphi')\) are glued together to form a single point of \( X_1 \). It follows that \( \rho \) induces a bijection \( \check{\rho}: X_1 \to \mathbb{T}^2 \) defined by the property that \( \check{\rho}(\check{P}) = \rho(P) \) for any \( P \in \bar{P} \).

The map \( \rho \) is continuous by the usual calculus arguments. Using the property that \( d(P, Q) \leq d(P, Q) \) (Lemma 5.2), it easily follows that \( \check{\rho} \) is continuous.

To prove that the inverse function \( \check{\rho}^{-1}: \mathbb{T}^2 \to \check{X}_1 \) is continuous at the point \( Q_0 \in \mathbb{T}^2 \), we need to distinguish cases according to the type of the point \( Q_0 \).

Consider the most complex case, where \( Q_0 = (-R + r, 0, 0) \) is the image under \( \check{\rho} \) of the point \( \check{P}_0 \in \check{X}_1 \) corresponding to the four vertices \((\pm \pi, \pm \pi)\) of \( X_1 \). If \( Q = (x, y, z) \in \mathbb{T}^2 \) is near \( Q_0 \), we can explicitly compute all \((\theta, \varphi)\) such that \( \rho(\theta, \varphi) = Q \). Indeed, \( \varphi = \pi - \arcsin \frac{y}{R + r \cos \varphi} \) if \( y > 0 \), \( \theta = -\pi - \arcsin \frac{y}{R + r \cos \varphi} \) if \( y < 0 \), and \( \varphi = \pm \pi \) if \( y = 0 \). Similarly, \( \theta = \pi - \arcsin \frac{y}{R + r \cos \varphi} \) if \( y > 0 \), \( \theta = -\pi - \arcsin \frac{y}{R + r \cos \varphi} \) if \( y < 0 \), and \( \theta = \pm \pi \) if \( y = 0 \). By continuity of the function \( \arcsin \), it follows that \((\theta, \varphi)\) will be arbitrarily close to one of the corners \((\pm \pi, \pm \pi)\) if \( Q = (x, y, z) \) is sufficiently close to \( Q_0 = (-R + r, 0, 0) \).

For \( \varepsilon > 0 \) small enough, Lemma 5.5 shows that the ball \( B_{\varepsilon}(\check{P}_0, \varepsilon) \) is just the image under the quotient map \( X_1 \to \check{X}_1 \) of the four quarter-disks of radius \( \varepsilon \) centered at the four vertices.
(±π, ±π) of X₁. By the observations above, there exists a δ > 0 such that, whenever Q ∈ T² is such that \( d_{\text{euc}}(Q, Q₀) < δ \), any (θ, φ) ∈ X₁ with \( ρ(θ, φ) = Q \) is within a distance < ε of one of the vertices (±π, ±π). Since \( \bar{ρ}^{-1}(Q) \) is the image in \( \bar{X}_1 \) of any (θ, φ) ∈ X₁ with \( ρ(θ, φ) = Q \), we conclude that \( d(\bar{ρ}^{-1}(Q), P₀) < ε \).

Since \( P₀ = \bar{ρ}^{-1}(Q₀) \), this proves that \( \bar{ρ}^{-1} \) is continuous at \( Q₀ = (−R + r, 0, 0) \).

The continuity at the other \( Q₀ ∈ T² \) is proved by a similar case-by-case analysis. □

6.1.2. Euclidean Klein Bottles. Given a rectangle \( X₃ = [a, b] × [c, d] \), we can also glue its sides together using different gluing maps. For instance, we can still glue the bottom edge \( E₁ = [a, b] × \{c\} \) to the top edge \( E₂ = [a, b] × \{d\} \) by the isometry \( ϕ₁: [a, b] × \{c\} → [a, b] × \{d\} \) defined by \( ϕ₁(x, c) = (x, d) \), but glue the left edge \( E₃ = \{a\} × [c, d] \) to the right edge \( E₄ = \{b\} × [c, d] \) by \( ϕ₃: \{a\} × [c, d] → \{b\} × [c, d] \) defined by \( ϕ₃(a, y) = (b, d − y) \). Namely, the gluing map flips the left edge upside down before sending it to the right edge by a translation.

We can consider a variation of this example, by replacing the rectangle by a parallelogram \( X₂ \), and gluing again opposite sides by translations. As in the case of the rectangle, the four vertices of the parallelogram are glued to a single point. Because the angles of a euclidean parallelogram add up to \( 2π \), the angle condition of Theorem 5.4 is satisfied, and we conclude that the quotient metric space \((\bar{X}_2, \bar{d})\) is a euclidean surface.

The parallelogram \( X₂ \) can clearly be stretched to assume the shape of a rectangle, in such a way that the gluing data for \( X₂ \) gets transposed to the gluing data for \( X₁ \). See Exercise 6.1. It follows that the quotient surface \( \bar{X}_2 \) is again a torus.

Again, the four vertices of \( X₃ \) are glued together to form a single point of the quotient space \( \bar{X}_3 \). Since the angles of \( X₃ \) at these four vertices add up to \( 2π \), the combination of Theorems 5.3 and 5.4 shows that \((\bar{X}_3, \bar{d}_{X₃})\) is locally isometric to the euclidean plane \((\mathbb{R}², d_{\text{euc}})\).

To understand the global shape of \( \bar{X}_3 \), we first glue the bottom and top sides together, to form a cylinder as in the case of the torus. We then need to glue the left side of the cylinder to the right side by a translation followed by a flip. This time, the difficulty of physically realizing this in 3-dimensional space goes well beyond the need for stretching the paper.
can actually be shown to be impossible to realize, in the sense that there is no subset of $\mathbb{R}^3$ which is homeomorphic to $\bar{X}_3$.

The right hand side of Figure 6.3 offers an approximation, where the surface crosses itself along a closed curve. Each point of this self-intersection curve has to be understood as corresponding to two points of the surface $\bar{X}_3$. Introducing an additional space dimension, this picture can also be used to represent an object in 4–dimensional space. This is similar to the way a figure eight $\infty$ in the plane can be deformed to a curve $\infty$ with no self-intersection in 3–dimensional space, by pushing parts of the figure eight up and down near the point where it crosses itself. In the same way, the object represented on the right-hand side of Figure 6.3 can be deformed to a subset of the 4–dimensional space $\mathbb{R}^4$ that is homeomorphic to $\bar{X}_3$.

The surface $\bar{X}_3$ is a **Klein bottle**.

The Klein bottle was introduced in 1882 by Felix Klein (1849 - 1925), as an example of pathological surface. The “bottle” terminology is usually understood to reflect the fact that a Klein bottle can we obtained from a regular wine bottle by stretching its neck and connecting it to the base after passing inside of the bottle. Another (unverified, and not incompatible with the previous one) interpretation claims that it comes from a bad pun, or a bad translation from the German, in which the *Kleinsche Fläche* (Klein surface) became the *Kleinsche Flasche* (Klein bottle). The latter version probably provides a better story, whereas the first one makes for better pictures. This second point is well illustrated by the
6.1. SOME EUCLIDEAN SURFACES

physical glass model of Figure 6.4, taken from the web site www.kleinbottle.com which offers for sale many Klein bottle shaped products.

6.1.3. Gluing opposite sides of a hexagon. Let us now go beyond quadrilaterals, and consider a hexagon \( X_4 \) where we glue opposite edges together, as indicated on Figure 6.5; see also Figure 5.3 and 5.6. The vertices of \( X_4 \) project to two points of the quotient space \( \overline{X}_4 \), each corresponding to three vertices of \( X_4 \). More precisely, if we label the vertices as \( P_1, P_2, \ldots, P_6 \) in this order as one goes around the hexagon, the odd vertices \( P_1, P_3 \) and \( P_5 \) are glued together to form one point of \( \overline{X}_4 \), and the even vertices \( P_2, P_4 \) and \( P_6 \) form another point of \( X_4 \). We consequently need the hexagon \( X_4 \) to satisfy the following two conditions:

1. opposite edges have the same length;
2. the angles of \( X_4 \) at the odd vertices \( P_1, P_3, P_5 \) add up to \( 2\pi \).

Recall that the sum of the angles of a euclidean hexagon is always equal to \( 4\pi \), so that Condition (2) is equivalent to the property that the angles of \( X_4 \) at its even vertices \( P_2, P_4, P_6 \) add up to \( 2\pi \). A little exercise in elementary euclidean geometry shows that, in a hexagon satisfying the above Conditions (1) and (2), opposite edges are necessarily parallel; see Exercise 6.7.

![Figure 6.5. Gluing opposite sides of a hexagon](image)

If the hexagon \( X_4 \) satisfies Conditions (1) and (2), we can again apply Theorems 5.3 and 5.4 to show that the quotient metric space \((\overline{X}_4, \overline{d})\) is a euclidean surface.

To understand the global shape of \( \overline{X}_4 \) up to homeomorphism, we can consider the diagonals \( P_1P_3 \) and \( P_2P_4 \) of the hexagon \( X_4 \), as in Figure 6.5. These two diagonals cut the hexagon into three pieces, the parallelogram \( P_1P_2P_4P_5 \) and the two triangles \( P_2P_3P_4 \) and \( P_3P_5P_1 \). Gluing the edges \( P_1P_2 \) and \( P_4P_5 \) of the parallelogram \( P_1P_2P_4P_5 \) provides a cylinder, whose boundary consists of the two images of the diagonals \( P_1P_3 \) and \( P_2P_4 \). Similarly, gluing the two triangles \( P_2P_3P_4 \) and \( P_3P_5P_1 \) together by identifying the edge \( P_2P_3 \) to the edge \( P_5P_6 \) and the edge \( P_3P_4 \) to \( P_6P_1 \) gives another cylinder, whose boundary again corresponds to the images of the diagonals \( P_1P_3 \) and \( P_2P_4 \). See the right hand side of Figure 6.5. This proves that splitting the quotient space \( \overline{X}_4 \) along the images of the diagonals \( P_1P_3 \) and \( P_2P_4 \) gives two cylinders. In particular, \( \overline{X}_4 \) can be recovered from these two cylinders by gluing them back together according to the pattern described on the right of Figure 6.5. It easily follows that \( \overline{X}_4 \) is homeomorphic to the torus.

As announced in an earlier disclaimer, this discussion of the construction of a homeomorphism from the quotient space \( \overline{X}_4 \) to the torus is somewhat informal. However, with a little reflection, you should be able to convince yourself that this description could be made completely rigorous if needed. The same will apply to other informal descriptions of homeomorphisms later on.
6.2. The surface of genus 2

We saw in the last section that, if we start with a euclidean rectangle or parallelogram, and if we glue opposite edges by a translation, we obtain a euclidean torus. Similarly, for a euclidean hexagon satisfying appropriate conditions on its edge lengths and angles, we found out that gluing opposite edges again yields a euclidean torus.

We can go one step further, and glue opposite sides of an octagon $X$, as on the left side of Figure 6.6.

![Figure 6.6. Gluing opposite edges of an octagon](image)

We claim that, the quotient space $\bar{X}$ is homeomorphic to the surface of genus 2 represented on the right of Figure 6.6, namely a torus with 2 handles.

To see this, one can cut out a smaller octagon from $X$ as indicated on Figure 6.6. This smaller octagon $X_1$ can be seen as a rectangle whose corners have been cut off. Gluing opposite edges of this rectangle, we see that the image $\bar{X}_1$ in $\bar{X}$ is just a torus from which a square (corresponding to the triangles removed from the rectangle) has been removed. See Figure 6.7.

![Figure 6.7. One half of Figure 6.6](image)

It remains to consider the four strips forming the complement $X_2$ of $X_1$ in $X$. Gluing these four strips together along their short edges gives a big square minus a smaller square, namely some kind of square annulus as on the left picture of Figure 6.8. Flipping this annulus inside out as in the middle picture of Figure 6.8, and then gluing the outside sides, we see that the image $\bar{X}_2$ of $X_2$ in the quotient space $\bar{X}$ is again a torus minus a square.

![Figure 6.8. The other half of Figure 6.6](image)

Finally, the quotient space $\bar{X}$ is obtained by gluing the two surfaces $\bar{X}_1$ and $\bar{X}_2$ along their boundaries, which gives the surface of genus 2 of Figure 6.6.
Let us now try to put a euclidean metric on the quotient space $\bar{X}$. The 8 vertices of the octagon $X$ are glued together to form a single point of $\bar{X}$. If we want to apply Theorems 5.3 and 5.4, we consequently need to use a euclidean octagon where opposite edges have the same length, and where the sum of the angles at the vertices is equal to $2\pi$. Unfortunately, in euclidean geometry, the angles of an octagon add up to $6\pi$!

It consequently seems impossible to put a euclidean metric on $\bar{X}$. (It can be proved that this is indeed the case, and that the surface of genus 2 admits no euclidean metric; see Exercise 6.16). However, hyperbolic geometry will provide us with a suitable octagon.

The first step is the following.

**Lemma 6.2.** In the hyperbolic plane $\mathbb{H}^2$, there exists a triangle $T$ with angles $\frac{\pi}{2}$, $\frac{\pi}{8}$ and $\frac{\pi}{8}$.

See Proposition 6.13, and Exercise 6.15, for a more general construction of hyperbolic triangles with prescribed angles.

**Proof.** We will actually use euclidean geometry to construct this hyperbolic triangle.

We begin with the hyperbolic geodesic $g$ with end points 0 and $\infty$. Namely $g$ is the vertical half-line beginning at 0. Then consider the complete geodesic $h$ that is orthogonal to $g$ at the point $i$. Namely, $h$ is the euclidean semi-circle of radius 1 centered at 0. We are looking for a third geodesic $k$ which makes an angle of $\frac{\pi}{8}$ with both $g$ and $h$.

For every $y \leq 1$, let $k_y$ be the complete geodesic that passes through the point $iy$ and makes an angle of $\frac{\pi}{8}$ with $g$. Namely, $k_y$ is a euclidean semi-circle of radius $y \csc \frac{\pi}{8}$ centered at the point $y \cot \frac{\pi}{8}$, and consequently meets $g$ as long as $(\sin \frac{\pi}{8}) / (1 + \cos \frac{\pi}{8}) < y \leq 1$. See Figure 6.9.

![Figure 6.9. A hyperbolic triangle with angles $\frac{\pi}{2}$, $\frac{\pi}{8}$ and $\frac{\pi}{8}$](image)

The angle $\alpha_y$ between $k_y$ and $g$ at their intersection point depends continuously on $y$. It is equal to $\frac{3\pi}{8}$ when $y = 1$, and approaches 0 as $y$ tends to $(\sin \frac{\pi}{8}) / (1 + \cos \frac{\pi}{8})$. By the Intermediate Value Theorem, there consequently exists a value of $y$ for which $\alpha_y = \frac{\pi}{8}$. By applying the Cosine Formula to the triangle formed by the intersection point of $g$ and $k_y$ and by the centers of these euclidean semi-circles, one could actually find an explicit formula for this $y$, but this is not necessary.

For this value of $y$, the hyperbolic geodesics $g$, $h$ and $k_y$ delimit the hyperbolic triangle $T$ required. □

**Lemma 6.3.** In the hyperbolic plane $\mathbb{H}^2$, there exists an octagon $X$ where all edges have the same length and where all angles are equal to $\frac{\pi}{4}$.
GLUING EXAMPLES

Proof. Let \( T \) be the triangle provided by Lemma 6.2. List its vertices as \( P_0, P_1 \) and \( P_2 \) in such a way that \( P_0 \) is the vertex with angle \( \frac{\pi}{2} \).

We start with 16 isometric copies \( T_1, T_2, \ldots, T_{16} \) of \( T \). Namely, the \( T_i \) are hyperbolic triangles for which there exists isometries \( \varphi_i : \mathbb{H}^2 \to \mathbb{H}^2 \) such that \( T_i = \varphi_i(T) \).

![Figure 6.10. A hyperbolic octagon with all angles equal to \( \frac{\pi}{4} \)](image)

Pick an arbitrary point \( Q \in \mathbb{H}^2 \). We can choose the isometries \( \varphi_i \) so that \( \varphi_i(P_2) = Q \) for every \( i \). In addition, using Lemma 5.8, we can arrange that:

1. If \( i \) is even, \( \varphi_i \) and \( \varphi_{i-1} \) send the edge \( P_0P_2 \) to the same geodesic, so that \( T_i \) and \( T_{i-1} \) have this edge \( \varphi_i(P_0P_2) = \varphi_{i-1}(P_0P_2) \) in common;
2. If \( i > 1 \) is odd, \( \varphi_i \) and \( \varphi_{i-1} \) send the edge \( P_1P_2 \) to the same geodesic, so that \( T_i \) and \( T_{i-1} \) have this edge \( \varphi_i(P_1P_2) = \varphi_{i-1}(P_1P_2) \) in common;
3. For every \( i > 1 \), the triangles \( T_i \) and \( T_{i-1} \) sit on opposite sides of their common edge.

Since \( 16 \frac{\pi}{8} = 2\pi \), the \( T_i \) fit nicely together around the point \( Q \), and the last edge of \( T_{16} \) comes back to match the first edge of \( T_1 \). In particular, \( \varphi_{16}(P_1P_2) = \varphi_1(P_1P_2) \). See Figure 6.10.

When \( i \) is even, the two geodesic arcs \( \varphi_i(P_0P_1) = \varphi_{i-1}(P_0P_1) \) meet at \( \varphi_i(P_0) = \varphi_{i-1}(P_0) \), and make an angle of \( \frac{\pi}{2} + \frac{\pi}{2} = \pi \) at that point. It follows that the union of these two geodesic arcs forms a single geodesic arc.

Therefore, the union \( X \) of the 16 triangles \( T_i \) is an octagon in the hyperbolic plane \( \mathbb{H}^2 \). Its angles are all equal to \( 2\frac{\pi}{8} = \frac{\pi}{4} \). Its edges all have the same length, namely twice the length of the edge \( P_0P_1 \) of the original triangle \( T \).

The symmetries of this hyperbolic octagon are more apparent in Figure 6.11, which represents its image in the disk model for \( \mathbb{H}^2 \) introduced in Section 3.7, if we arrange that the center \( Q \) of \( X \) corresponds to the center \( O \) of \( \mathbb{B}^2 \).

Let \( X \) be the hyperbolic octagon provided by Lemma 6.3, and let \( \tilde{X} \) be the quotient space obtained by gluing together opposite sides of \( X \), as in Figure 6.6. Then Theorems 5.9 and 5.10 assert that the metric \( d_X = d_{hyp} \) induces a quotient metric \( \tilde{d}_X \) on this quotient space \( X \), and that the metric space \( (\tilde{X}, \tilde{d}_X) \) is locally isometric to the hyperbolic plane.

In particular, we have constructed a hyperbolic surface \( (\tilde{X}, \tilde{d}_X) \) which is homeomorphic to the surface of genus 2.

6.3. The projective plane

We now construct a spherical surface, which is different from the sphere \( \mathbb{S}^2 \).
Let $X$ be a hemisphere in $S^2$. To turn $X$ into a polygon, pick two antipodal points $P_1$ and $P_2 = -P_1$ on the great circle $C$ delimiting $X$, which will be the vertices of the polygon. These two vertices split $C$ into two edges $E_1$ and $E_2$. We now have a spherical polygon $X$, which we endow with the path metric $d_X$. Note that $X$ is convex, so that $d_X$ is just the restriction of the spherical metric $d_{sph}$.

Now, glue $E_1$ to $E_2$ by the antipode map $\varphi_1: E_1 \to E_2$ defined by $\varphi_1(P) = -P$. This gluing data defines a quotient space $(\tilde{X}, \tilde{d}_X)$.

The polygon $X$ and its gluing data are represented in Figure 6.12.

The angles of $X$ at $P_1$ and $P_2$ are both equal to $\pi$, and consequently add up to $2\pi$. We can consequently apply Theorems 5.12 and 5.13, and show that the quotient space $(\tilde{X}, \tilde{d}_X)$ is locally isometric to the sphere $(S^2, d_{sph})$. This quotient space $(\tilde{X}, \tilde{d}_X)$ is called the projective plane.

In Exercise 6.10, we show that the projective plane can also be interpreted as the space of lines passing through the origin in the 3–dimensional space $\mathbb{R}^3$.

6.4. The cylinder and the Möbius strip

We now consider unbounded polygons, and the surfaces obtained by gluing their edges together.

The simplest case is that of an infinite strip where the two edges are glued together, which provides a cylinder or a Möbius strip. These examples may appear somewhat trivial at first, but they already display many features that we will encounter in more complicated surfaces.

6.4.1. Euclidean cylinders and Möbius strips. We can begin with an infinite strip $X_1$ in the euclidean plane $\mathbb{R}^2$, bounded by two parallel lines $E_1$ and $E_2$. Orient $E_1$ and $E_2$ in
the same direction, and glue them by an isometry \( \varphi_1: E_1 \to E_2 \) respecting these orientations. Because there are no vertices on \( E_1 \) and \( E_2 \), this is a situation where the gluing map \( \varphi_1 \) is not uniquely determined by these properties. Indeed, there are many possible choices for \( \varphi_1 \), all differing from each other by composition with a translation of \( \mathbb{R}^2 \) parallel to \( E_1 \) and \( E_2 \).

Pick any such gluing map \( \varphi_1: E_1 \to E_2 \), and consider the corresponding quotient space \((\bar{X}_1, \bar{d}_{X_1})\). Since \( X_1 \) has no vertex, the angle hypothesis of Theorem 5.4 is automatically satisfied. Therefore Theorems 5.3 and 5.4 show that \((\bar{X}_1, \bar{d}_{X_1})\) is a euclidean surface.

This euclidean surface is easily seen to be homeomorphic to the cylinder.

![Figure 6.13. A euclidean cylinder](image)

For another vertical strip \( X_2 \) in \( \mathbb{R}^2 \) bounded by parallel lines \( E_1 \) and \( E_2 \), we can orient \( E_1 \) and \( E_2 \) so that they now point in opposite directions, and choose a gluing map \( \varphi: E_1 \to E_2 \) respecting these orientations. Again, there are many different choices for this gluing map, differing by composition with translations parallel to \( E_1 \) and \( E_2 \).

For each choice of such a gluing map \( \varphi_1 \), another application of Theorems 5.3 and 5.4 shows that the corresponding quotient space \((\bar{X}_2, \bar{d}_{X_2})\) is a euclidean surface.

The topology of the quotient space \((\bar{X}_2, \bar{d}_{X_2})\) is now very different. Indeed, this space is homeomorphic to the famous Möbius strip.

![Figure 6.14. A euclidean Möbius strip](image)

The Möbius strip is named after August Möbius (1790–1868), who conceived of this non-orientable surface in 1858 while working on geometric properties of polyhedra. He never published this work, which was only discovered after his death. Credit for the discovery of the Möbius strip should probably go to Johann Benedict Listing (1808–1882) instead, who independently described the Möbius strip in 1858. Incidentally, Listing made another
important contribution to the themes of this monograph. He coined the word “topology” (or Topologie in the German original) in an 1836 letter, and the first printed occurrence of this term appears in his book Vorstudien zur Topologie, published in 1847.

6.4.2. Hyperbolic cylinders. We can make completely analogous constructions in hyperbolic geometry by replacing the euclidean strip with a strip \( X_3 \) in the hyperbolic plane \( \mathbb{H}^2 \), bounded by two disjoint complete geodesics \( E_1 \) and \( E_2 \). However, there are essentially two different shapes for such an infinite strip in \( \mathbb{H}^2 \).

First, the end points of \( E_1 \) and \( E_2 \) on \( \mathbb{R} \cup \{ \infty \} \) may all be distinct. By an easy algebraic exercise with linear fractional maps, there exists an isometry of \( \mathbb{H}^2 \) sending \( E_1 \) to the complete geodesic with end points \( -1 \) and sending \( E_2 \) to the complete geodesic with end points \( -a \), for some \( a < 1 \). We can consequently assume, without loss of generality, that \( E_1 \) goes from \(-1\) to \(+1\) and that \( E_2 \) goes from \(-a\) to \(+a\).

To glue \( E_1 \) to \( E_2 \) we need an isometry \( \varphi_1 \) sending \( E_1 \) to \( E_2 \). Among such isometries of \( \mathbb{H}^2 \), the simplest one is the homothety defined by \( \varphi_1(z) = az \). Choose this specific isometry as a gluing map, and let \( (\bar{X}_3, \bar{d}_{X_3}) \) be the corresponding quotient metric space. The combination of Theorems 5.9 and 5.10 shows that \( (\bar{X}_3, \bar{d}_{X_3}) \) is locally isometric to the hyperbolic plane \( (\mathbb{H}^2, d_{\text{hyp}}) \), namely is a hyperbolic surface. We are now in the situation of Figure 6.15.

This surface is easily shown to be homeomorphic to the cylinder. However, its geometry is very different from that of a euclidean cylinder.

Indeed, in the euclidean case, assume that \( X_1 \) is the vertical strip \( \{(x, y) \in \mathbb{R}^2; 0 \leq x \leq a \} \) and that we glue \( E_1 = \{(x, y) \in \mathbb{R}^2; x = 0\} \) to \( E_2 = \{(x, y) \in \mathbb{R}^2; x = a\} \) by the horizontal translation \( \varphi_1: (x, y) \mapsto (x + a, y) \). The quotient space \( (\bar{X}_1, \bar{d}_{X_1}) \) can then be decomposed as a union of closed curves \( \gamma_t \) where, for each \( t \), \( \gamma_t \) is the image in \( \bar{X}_1 \) of the horizontal line segment \( \{(x, y) \in \mathbb{R}^2; 0 \leq t \leq a, y = t\} \). These closed curves all have length \( a \), and the set of points at distance \( \delta \) from the central curve \( \gamma_0 \) exactly consists of the two curves \( \gamma_\delta \cup \gamma_{-\delta} \). In particular, the curves \( \gamma_t \) show that the “width” of the euclidean cylinder \( \bar{X}_1 \) is the same at every point.

In the hyperbolic case, where \( X_3 \) is the strip in \( \mathbb{H}^2 \) delimited by the geodesics \( E_1 \) and \( E_2 \) respectively going from \(-1\) to \(+1\) and from \(-a\) to \(+a\), we glued \( E_1 \) to \( E_2 \) by the homothety \( \varphi_1(z) = az \). For every \( \theta \) with \(-\frac{\pi}{2} < \theta < \frac{\pi}{2}\), we have a closed curve \( \gamma_\theta \) in the quotient space \( \bar{X}_3 \), image of the euclidean line segment consisting of all \( z = re^{i(\frac{\pi}{2}-\theta)} \) with \( a \leq r \leq 1 \).
The closed curve $\gamma_0$ is geodesic because, for each $\bar{P} \in \gamma_0$, there exists an isometry between a small ball $B_{d_{hyp}}(\bar{P}, \varepsilon)$ and a ball of radius $\varepsilon$ in $(\mathbb{H}^2, d_{hyp})$ sending $\gamma_0 \cap B_{d_{hyp}}(\bar{P}, \varepsilon)$ to a geodesic arc of $\mathbb{H}^2$. The only point $\bar{P}$ where this requires a little checking is when $\bar{P}$ corresponds to the two points $i$ and $ai \in X_3$, in which case the local isometry provided by the proof of Theorem 5.10 is easily seen to satisfy this property. The curves $\gamma_\theta$ with $\theta \neq 0$ are never geodesic.

A few rather immediate computations of hyperbolic lengths show the following:

1. The curve $\gamma_\theta$ has hyperbolic length $\int_1^a \frac{1}{r \cos \theta} dr = \log \sec \theta$.

2. Every point $z = re^{i(\frac{\pi}{2} - \theta)}$ of $\gamma_\theta$ is at distance $\int_0^{[\theta]} \frac{r}{r \cos t} dt = \log(\sec \theta + \tan |\theta|)$ from the curve $\gamma_0$.

Compare Exercise 3.5 for the proof of (2).

Therefore, for every $\delta > 0$, the set of points of $\bar{X}_3$ that are at distance $\delta$ from $\gamma_0$ consists of the two curves $\gamma_{\pm \theta}$ with $\log(\sec \theta + \tan |\theta|) = \delta$. By elementary trigonometry, this is equivalent to the property that $\sec \theta = \cosh \delta$. Therefore, by (1) above, the length of each of these two curves is equal to $\log a \cosh \delta$. In particular the width of the cylinder grows exponentially with the distance $\delta$ from $\gamma_0$, so that $\gamma_0$ forms some kind of ‘narrow waist’ for the hyperbolic cylinder $\bar{X}_3$.

The picture on the right-hand side of Figure 6.15 attempts to convey a sense of this exponential growth. This picture is necessarily imperfect. Indeed, as one goes towards one of the ends of the hyperbolic cylinder, its width grows faster than that of any surface of revolution in the 3–dimensional euclidean space $\mathbb{R}^3$.

In the hyperbolic plane $\mathbb{H}^2$, there is another type of infinite strip bounded by two complete geodesics $E_1$ and $E_2$, which occurs when $E_1$ and $E_2$ have one end point in common in $\mathbb{R} \cup \{\infty\}$. Applying an isometry of $\mathbb{H}^2$, we can assume without loss of generality that this common point is $\infty$, namely that $E_1$ and $E_2$ are both vertical half-lines. By a horizontal translation followed by a homothety, we can even arrange that $E_1$ is the vertical half-line with end points 0 and $\infty$, while $E_2$ goes from 1 to $\infty$. These two geodesics now delimit the strip $X_4 = \{z \in \mathbb{H}^2; 0 \leq \text{Re}(z) \leq 1\}$ in $\mathbb{H}^2$.

To glue the edges $E_1$ and $E_2$ together, the simplest gluing map $\varphi_1 : E_1 \to E_2$ is the horizontal translation defined by $\varphi_1(z) = z + 1$. Let $(\bar{X}_4, \bar{d}_{hyp})$ be the quotient space obtained
by performing this gluing operation on \( X_4 \). This quotient space is a hyperbolic surface by Theorems 5.9 and 5.10, and is easily seen to be homeomorphic to the cylinder. We should note that our choice of the gluing map \( \varphi_1 \) is here critical. Indeed, we will see in Section 7.7.1 that other choices lead to hyperbolic cylinders with very different geometric properties.

For every \( t \in \mathbb{R} \), let \( \gamma_t \) be the closed curve in the quotient space \( \bar{X}_4 \) that is the image of the horizontal line segment consisting of those \( z \in X_4 \) such that \( \text{Im}(z) = e^t \). By definition of the hyperbolic metric, it is immediate that \( \gamma_t \) has hyperbolic length \( e^{-t} \). Also, every point of \( \gamma_t \) is at hyperbolic distance \( |t| \) from the central curve \( \gamma_0 \). Therefore, the width of \( \bar{X}_4 \) grows exponentially towards one end of the cylinder, and decreases exponentially towards the other end.

Again, the right hand side of Figure 6.16 attempts to illustrate this behavior. As in the case of \( \bar{X}_3 \), this picture is necessarily imperfect for the end with exponential growth. Surprisingly enough, the end with exponential decay can be exactly represented as a surface of revolution in the 3–dimensional space \( \mathbb{R}^3 \).

![Figure 6.17. The pseudosphere](image)

More precisely, let \( X_4^+ \) denote the upper part of \( X_4 \) consisting of those \( z \in X_4 \) such that \( \text{Im}(z) \geq \frac{1}{2\pi} \), and let \( \bar{X}_4^+ \) be its image in \( \bar{X}_4 \).

In the \( xy \)-plane, consider the \textit{tractrix} parametrized by

\[
t \mapsto (t - \tanh t, \text{sech } t), \quad 0 \leq t < \infty,
\]

and let the \textit{pseudosphere} \( S \) be the surface of revolution in \( \mathbb{R}^3 \) obtained by revolving the tractrix about the \( x \)-axis. This surface is represented on Figure 6.17.

Endow the pseudosphere \( S \) with the metric \( d_S \) defined by the property that, for every \( P, Q \in S \), \( d_S(P, Q) \) is the infimum of the euclidean lengths \( l_{\text{euc}}(\gamma) \) of all piecewise differentiable curves \( \gamma \) contained in \( S \) and joining \( P \) to \( Q \).

As usual, we endow \( X_4^+ \) with the metric \( d_{X_4^+} \) for which the distance between two points is the infimum of the hyperbolic lengths of all curves in \( X_4^+ \) joining these two points. Because hyperbolic geodesics are euclidean semi-circles, it is relatively immediate that \( X_4^+ \) is convex, so that this metric \( d_{X_4^+} \) actually coincides with the restriction of the hyperbolic distance. However, it is convenient to keep a distinct notation because of (minor) subtleties with
quotient metrics; compare Lemma 6.8. Let \( \tilde{d}_{X^+_4} \) be the quotient metric induced by \( d_{X^+_4} \) on the quotient space \( \bar{X}^+_4 \).

**Proposition 6.4.** The metric space \((\bar{X}^+_4, \tilde{d}_{X^+_4})\) is isometric to the surface \((S, d_S)\).

**Proof.** The proof will take several steps, in part because of the definition of the quotient metric. It may perhaps be skipped on a first reading. However, you should at least have a glance at Lemma 6.5, which contains the key geometric idea of the proof.

Let \( \rho: X^+_4 \to S \) be the map which to \( z \in X^+_4 \) associates the point \( \rho(z) = (u, v, w) \) with

\[
\begin{align*}
u &= \log \left( 2\pi \text{Im}(z) + \sqrt{4\pi^2 \text{Im}(z)^2 - 1} \right) - \frac{\sqrt{4\pi^2 \text{Im}(z)^2 - 1}}{2\pi \text{Im}(z)} \\
v &= \frac{1}{2\pi \text{Im}(z)} \cos(2\pi \text{Re}(z)) \\
w &= \frac{1}{2\pi \text{Im}(z)} \sin(2\pi \text{Re}(z))
\end{align*}
\]

Geometrically, \( \rho(z) \) is better understood by setting

\[ t = \arccosh(2\pi \text{Im}(z)) = \log \left( 2\pi \text{Im}(z) + \sqrt{4\pi^2 \text{Im}(z)^2 - 1} \right). \]

Then, \( \rho(z) = (t - \tanh t, \text{sech} \ t \cos(2\pi \text{Re}(z)), \text{sech} \ t \sin(2\pi \text{Re}(z))) \) is obtained by rotating by an angle of \( 2\pi \text{Re}(z) \) about the \( x \)-axis the point of the tractrix corresponding to the parameter \( t \).

In particular, two points \( z, z' \in X^+_4 \) have the same image under \( \rho \) if and only if \( z' - z \) is an integer, namely if and only if \( z \) and \( z' \) are glued together to form a single point of the quotient space \( \bar{X}^+_4 \). As a consequence, \( \rho \) induces an injective map \( \bar{\rho}: \bar{X}^+_4 \to S \) defined by the property that \( \bar{\rho}(\bar{P}) \) is equal to \( \rho(P) \) for every \( P \in \bar{X}^+_4 \) and any element \( P \in \bar{P} \), namely for any point \( P \in X^+_4 \) that corresponds to \( \bar{P} \) in the quotient space \( \bar{X}^+_4 \).

The map \( \rho \) is surjective by definition of the pseudo-sphere \( S \). It follows that \( \bar{\rho}: \bar{X}^+_4 \to S \) is surjective, and is therefore bijective.

We will prove that \( \bar{\rho} \) is an isometry from \((\bar{X}^+_4, \tilde{d}_{X^+_4})\) to \((S, d_S)\). The key step is the following computation.

**Lemma 6.5.** The map \( \rho \) sends every curve \( \gamma \) in the half-strip \( X^+_4 \) to a curve in the surface \( S \) whose euclidean length \( \ell_{\text{euc}}(\rho(\gamma)) \) is equal to the hyperbolic length \( \ell_{\text{hyp}}(\gamma) \) of \( \gamma \).

**Proof.** Let us compute the differential map \( D_z \rho \). Remember that \( D_z \rho \) sends the vector \( \bar{v} = (a, b) \) to

\[
D_z \rho(\bar{v}) = \left( \frac{\partial v}{\partial x} a + \frac{\partial v}{\partial y} b, \frac{\partial w}{\partial x} a + \frac{\partial w}{\partial y} b, \frac{\partial w}{\partial x} a + \frac{\partial w}{\partial y} b \right)
\]

if we write \( z = x + iy \).

With \( t = \arccosh(2\pi y) \) as before, we find that \( \frac{\partial u}{\partial x} = 0, \frac{\partial u}{\partial y} = 2\pi \text{sech} \ t \tanh t, \frac{\partial v}{\partial x} = -2\pi \text{sech} \ t \sin(2\pi x), \frac{\partial v}{\partial x} = 2\pi \text{sech}^2 t \cos(2\pi x), \frac{\partial w}{\partial x} = 2\pi \text{sech} t \cos(2\pi x), \) and \( \frac{\partial w}{\partial x} = 2\pi \text{sech}^2 t \sin(2\pi x) \).

After simplifications,

\[
\|D_z \rho(\bar{v})\|_{\text{euc}} = 2\pi \text{sech} t \|\bar{v}\|_{\text{euc}} = \frac{1}{y} \|\bar{v}\|_{\text{euc}} = \|\bar{v}\|_{\text{hyp}},
\]
If $\gamma$ is a curve parametrized by $s \mapsto z(s)$, $s_1 \leq s \leq s_2$, its image $\rho(\gamma)$ is parametrized by $s \mapsto \rho(z(s))$, $s_1 \leq s \leq s_2$. Therefore,

$$\ell_{\text{euc}}(\varphi(\gamma)) = \int_{s_1}^{s_2} \|(\rho \circ z)'(s)\|_{\text{euc}} \, ds = \int_{s_1}^{s_2} \|D_z(z'(s))\|_{\text{euc}} \, ds$$

$$= \int_{s_1}^{s_2} \|z'(s)\|_{\text{hyp}} \, ds = \ell_{\text{hyp}}(\gamma).$$

\[\square\]

**Lemma 6.6.**

$$d_s(\bar{\rho}(\overline{P}), \bar{\rho}(\overline{Q})) \leq d_{X^+_4}(\overline{P}, \overline{Q})$$

for every $\overline{P}, \overline{Q} \in \overline{X^+_4}$.

**Proof.** Let the points $P = P_1, Q_1 \sim P_2, \ldots, Q_{n-1} \sim P_n, Q_n = Q$ form a discrete walk $w$ from $\overline{P}$ to $\overline{Q}$. The length of the discrete walk $w$ is $\ell(w) = \sum_{i=1}^{n} d_{X^+_4}(P_i, Q_i)$.

By convexity of $X^+_4$, there is a geodesic arc $\gamma_i$ joining $P_i$ to $Q_i$ whose hyperbolic length is equal to $d_{X^+_4}(P_i, Q_i)$. Lemma 6.5 shows that the image $\gamma'_i = \rho(\gamma_i)$ is a curve joining $\rho(P_i)$ to $\rho(Q_i) = \rho(P_{i+1})$ in $S$ whose euclidean length is equal to the hyperbolic length of $\gamma_i$. Chaining together these $\gamma'_i$ provides a curve $\gamma'$ joining $\rho(P) = \bar{\rho}(P)$ to $\rho(Q) = \bar{\rho}(Q)$ in $S$. Its euclidean length is

$$\ell_{\text{euc}}(\gamma') = \sum_{i=1}^{n} \ell_{\text{euc}}(\gamma_i) = \sum_{i=1}^{n} \ell_{\text{hyp}}(\gamma_i) = \sum_{i=1}^{n} d_{X^+_4}(P_i, Q_i) = \ell(w).$$

By definition of the metric $d_s$, this shows that $d_s(\bar{\rho}(\overline{P}), \bar{\rho}(\overline{Q})) \leq \ell(w)$. Since this holds for every discrete walk $w$ from $\overline{P}$ to $\overline{Q}$, we conclude that $d_s(\bar{\rho}(\overline{P}), \bar{\rho}(\overline{Q})) \leq d_{X^+_4}(\overline{P}, \overline{Q})$. \[\square\]

**Lemma 6.7.**

$$d_{X^+_4}(\overline{P}, \overline{Q}) \leq d_s(\bar{\rho}(\overline{P}), \bar{\rho}(\overline{Q}))$$

for every $\overline{P}, \overline{Q} \in \overline{X^+_4}$.

**Proof.** For a given $\varepsilon > 0$, there exists a piecewise differentiable curve $\gamma$ going from $\bar{\rho}(\overline{P})$ to $\bar{\rho}(\overline{Q})$ in $S$ whose length is sufficiently close to $d_s(\bar{\rho}(\overline{P}), \bar{\rho}(\overline{Q}))$ that

$$d_s(\bar{\rho}(\overline{P}), \bar{\rho}(\overline{Q})) \leq \ell_{\text{euc}}(\gamma) \leq d_s(\bar{\rho}(\overline{P}), \bar{\rho}(\overline{Q})) + \varepsilon.$$

We want to decompose $\gamma$ into pieces coming from $X^+_4$. For this, we use the following estimate.

In the surface $S$, consider the tractrix $T$ parametrized by $t \mapsto (t - \tanh t, t, 0)$. If $\alpha$ is a curve in $S$ whose end points $P'$ and $Q'$ are both in $T$, and if $\beta$ is the portion of $T$ going from $P'$ to $Q'$, parametrize $\alpha$ by $s \mapsto (x(s), y(s), z(s))$, $a \leq s \leq b$, with

\[
\begin{align*}
  x(s) &= t(s) - \tanh t(s) \\
  y(s) &= \sech t(s) \cos \theta(s) \\
  z(s) &= \sech t(s) \sin \theta(s)
\end{align*}
\]

where $\theta(s)$ is chosen so that $x(s)$ is the portion of $T$ from $P'$ to $Q'$.
for some functions \( s \mapsto t(s) \) and \( s \mapsto \theta(s) \). The curve \( \beta \) has a similar parametrization, where \( \theta(s) \) is constant equal to 0. An immediate computation then yields

\[
\ell_{\text{euc}}(\alpha) = \int_a^b \sqrt{x'(s)^2 + y'(s)^2 + z'(s)^2} \, ds \\
= \int_a^b \sqrt{t'(s)^2 \tanh^2 t(s) + \theta'(s)^2 \operatorname{sech}^2 t(s)} \, ds \\
\geq \int_a^b t'(s) \tanh t(s) \, ds = \ell_{\text{euc}}(\beta).
\]

As a consequence, we can arrange that the intersection of \( \gamma \) with \( T \) consists of a single curve contained in \( T \), possibly empty or reduced to a single point. Indeed, if \( P' \) and \( Q' \) are the first and last points where \( \gamma \) meets \( T \), we can replace the part \( \alpha \) of \( \gamma \) going from \( P' \) to \( Q' \) by the part \( \beta \) of \( T \) joining \( P' \) to \( Q' \) without increasing its length.

We are now ready to conclude. The key observation is that \( \gamma \) and \( \gamma' \) have the same image under \( \rho \) if and only if they are glued together in the quotient space \( X_4^+ \). By convexity of \( X_4 \), we can arrange that the intersection of \( \gamma \) and \( \gamma' \) is disjoint from any \( X_4^+ \). Consequently, if \( P_1 \) and \( Q_i \) denote the initial and end points of each \( \gamma_i \), we have that \( P \sim P_1, Q_i \sim P_2, Q_2 \sim P_3 \) and \( Q_3 \sim Q \). In particular, the \( P_i \) and \( Q_i \) form a discrete walk \( w \) from \( P \) to \( Q \) whose length is

\[
\ell_{d_{X_4^+}}(w) = \sum_{i=1}^3 d_{X_4^+}(P_i, Q_i) \leq \sum_{i=1}^3 \ell_{\text{hyp}}(\gamma_i') = \sum_{i=1}^3 \ell_{\text{euc}}(\gamma_i) = \ell_{\text{euc}}(\gamma)
\]

since the euclidean length of \( \gamma_i \) is equal to the hyperbolic length of \( \gamma_i' \) by Lemma 6.5. Therefore,

\[
\tilde{d}_{X_4^+}(P, Q) \leq d_{X_4^+}(P, Q) \leq \ell_{d_{X_4^+}}(w) \leq \ell_{\text{euc}}(\gamma) \leq d_S(\rho(P), \rho(Q)) + \varepsilon.
\]

We have now proved that \( \tilde{d}_{X_4^+}(P, Q) \leq d_S(\rho(P), \rho(Q)) + \varepsilon \) in both cases, and this for every \( \varepsilon > 0 \). Therefore, \( \tilde{d}_{X_4^+}(P, Q) \leq d_S(\rho(P), \rho(Q)) \) as requested.

The combination of Lemmas 6.6 and 6.7 shows that \( d_S(\rho(P), \rho(Q)) = \tilde{d}_{X_4^+}(P, Q) \) for every \( P, Q \in X_4^+, \) namely that \( \rho \) is an isometry from \( (X_4^+, \tilde{d}_{X_4^+}) \) to \( (S, d_S) \). This completes the proof of Proposition 6.4.

We conclude our discussion of Proposition 6.4 by addressing a little subtlety. By convexity of \( X_4 \) and \( X_4^+ \), \( d_{X_4^+}(P, Q) = d_{X_4}(P, Q) = d_{\text{hyp}}(P, Q) \) for every \( P \) and \( Q \in X_4^+ \). However,
in the quotient space $\tilde{X}_4^+$, there might conceivably be a difference between the quotient metric $\bar{d}_{X_4^+}$ and the restriction of the quotient metric $\bar{d}_{X_4}$ of $\tilde{X}_4$. Indeed, the definition of the quotient metric $\bar{d}_{X_4^+}$ involves discrete walks valued in $X_4$, whereas $\bar{d}_{X_4}$ is defined using discrete walks which are constrained to $X_4^+$. Compare our discussion of Lemma 6.12 in the next section.

In the specific case under consideration, it turns out that we do not need to worry about this distinction:

**Lemma 6.8.** On the subspace $X_4^+$ of $\tilde{X}_4$, the two metrics $\bar{d}_{X_4^+}$ and $\bar{d}_{X_4}$ coincide.

**Proof.** Because $X_4^+$ is contained in $X_4$ and because the metrics $d_{X_4}$ and $d_{X_4^+}$ coincide on $X_4^+$, every discrete walk $w$ valued in $X_4^+$ is also valued in $X_4$, and the two lengths $\ell_{d_{X_4^+}}(w)$ and $\ell_{d_{X_4}}(w)$ coincide. It follows that $\bar{d}_{X_4^+}(P, Q) \leq \bar{d}_{X_4}(P, Q)$ for every $P, Q \in X_4^+$.

Conversely, let $P = P_1, Q_1 \sim P_2, \ldots, Q_{n-1} \sim P_n, Q_n = Q$ be a discrete walk $w$ valued in $X_4$ where $P$ and $Q$ both belong to the upper half-strip $X_4^+$. For each $P_i$, let $P_i'$ be equal to $P_i$ if $P_i$ is in $X_4^+$, and otherwise let $P_i' = \text{Re}(P_i) + \frac{1}{2\pi}i$ be the point of the boundary of $X_4^+$ that sits right above $P_i$. Similarly define $Q_i'$ to be $Q_i$ if $Q \in X_4^+$, and $\text{Re}(Q_i) + \frac{1}{2\pi}i$ if $Q \not\in X_4^+$.

We claim that $d_{\text{hyp}}(P_i', Q_i') \leq d_{\text{hyp}}(P_i, Q_i)$. Indeed, if $\gamma$ is a curve joining $P_i$ to $Q_i$ in $X_4$, let $\gamma'$ be obtained from $\gamma$ by replacing each piece that lies below the line of equation $\text{Im}(z) = \frac{1}{2\pi}$ with the line segment that sits on the line right above that piece. From the definition of hyperbolic length, it is immediate that $\ell_{d_{\text{hyp}}}(\gamma') \leq \ell_{d_{\text{hyp}}}(\gamma)$. Considering all such curves $\gamma$ and $\gamma'$, it follows that $d_{\text{hyp}}(P_i', Q_i') \leq d_{\text{hyp}}(P_i, Q_i)$.

Since $P'_1 = P_1$ and $P'_n = P_n$, we now have a discrete walk $P = P'_1, Q'_1 \sim P'_2, \ldots, Q'_{n-1} \sim P'_n, Q'_n = Q$ which is valued in $X_4^+$. Since $d_{\text{hyp}}(P'_i, Q'_i) \leq d_{\text{hyp}}(P_i, Q_i)$, this new walk $w'$ has length $\ell_{d_{X_4^+}}(w') \leq \ell_{d_{X_4}}(w)$. As a consequence, $\ell_{d_{X_4}}(w) \geq \bar{d}_{X_4^+}(P, Q)$.

Considering all such walks $w$, we conclude that $\bar{d}_{X_4^+}(P, Q) \leq \bar{d}_{X_4}(P, Q)$. Therefore, $\bar{d}_{X_4^+}(P, Q) = \bar{d}_{X_4}(P, Q)$ for every $P, Q \in \tilde{X}_4^+$. □

### 6.5. The once-punctured torus

The **once-punctured torus** is obtained by removing one point from the torus. To explain the terminology, think of what happens to the inner tube of a tire as one drives over a nail. If we describe the torus as a square with opposite edges glued together, we can assume that the point removed is the point corresponding to the four vertices of the square.

This surface of course admits a euclidean metric, by restriction of a euclidean metric on the torus. However, we will see in the next Chapter 7 that such a metric is not complete (see definition in Section 7.2), and that complete metrics are more desirable. Our goal is to construct on the once-punctured torus a hyperbolic metric, which we will later on prove to be complete.

This example will turn out to be very important. In particular, it will accompany much of our discussion in Chapters 8, ?? and ??.

Consider the hyperbolic polygon $X$ described in Figure 6.18. Namely, $X$ is the region in the hyperbolic plane $\mathbb{H}^2$ bounded by the four complete geodesics $E_1$, $E_2$, $E_3$ and $E_4$ where $E_1$ joins $-1$ to $\infty$, $E_2$ joins $0$ to $1$, $E_3$ joins $1$ to $\infty$, and $E_4$ joins $0$ to $-1$. from Page 76
As such, $X$ is a “quadrilateral” except that its vertices are in $\mathbb{R} \cup \{\infty\}$, namely at infinity of $\mathbb{H}^2$. As a subset of $\mathbb{H}^2$, $X$ is therefore a quadrilateral with its four vertices removed.

In a hyperbolic polygon of this type, where the vertices are at infinity of $\mathbb{H}^2$ in $\mathbb{R} \cup \{\infty\}$ (and consequently are not really vertices of the polygon in $\mathbb{H}^2$), we say that the vertices are ideal. If the vertices of the polygon are all ideal, and if the polygon touches $\mathbb{R} \cup \{\infty\}$ only at these vertices, we say that it is an ideal polygon.

With this terminology, $X$ is now an ideal quadrilateral in $\mathbb{H}^2$.

Glue together opposite edges of $X$, while respecting the orientations indicated on Figure 6.18. Because these geodesics do not have any end points in $\mathbb{H}^2$, this is a situation where there are many possible isometric gluings. In particular, the gluing data is not completely determined by the picture. We consequently need to be more specific.

To glue the edge $E_1$ to $E_2$, we need a hyperbolic isometry $\varphi_1$ sending $-1$ to $0$, and $\infty$ to $1$. The simplest one is the linear fractional map

$$\varphi_1(z) = \frac{z + 1}{z + 2}.$$

Similarly, we can glue $E_3$ to $E_4$ by the hyperbolic isometry

$$\varphi_3(z) = \frac{z - 1}{-z + 2}.$$

As usual, define $\varphi_2 = \varphi_1^{-1}$ and $\varphi_4 = \varphi_3^{-1}$.

Let $(\tilde{X}, \tilde{d}_X)$ be the quotient metric space obtained from $(X, d_X)$ by performing these edge gluings. Note that $X$ is convex, so that the metric $d_X$ is just the restriction of the hyperbolic metric $d_{\text{hyp}}$.

Since $X$ has no vertices in $\mathbb{H}^2$, there is nothing to be checked and Theorem 5.10 shows that $(\tilde{X}, \tilde{d}_X)$ is a hyperbolic surface. From the description of $X$ as a quadrilateral with its vertices removed, we see that $\tilde{X}$ is (homeomorphic to) a once-punctured torus.

We want to better understand the metric $d_X$ near the puncture.

For $a > 1$ consider the horizontal line $L_a$ of equation $\text{Im}(z) = a$.

By Proposition 3.18, linear fractional maps send circles to circles in $\hat{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$. Consequently, $\varphi_1$ sends $L_a \cup \{\infty\}$ to a circle $C_1$ passing through $\varphi_1(\infty) = 1$. Since $\varphi_1$ also sends the half-plane $\mathbb{H}^2$ to itself, this circle must be tangent to the real line at $1$. The circle $C_1$ also contains the image $\varphi_1(-2 + ai) = 1 + \frac{1}{a}i$ of the point $-2 + ai \in L_a$. It follows that $C_1$ is the euclidean circle of radius $\frac{1}{2a}$ centered at $1 + \frac{1}{2a}i$, and $\varphi_1(L_a) = C_1 - \{1\}$. 
Similarly, $\varphi_3(L_a)$ is contained in a circle $C_{-1}$ tangent to the real line at $\varphi_3(\infty) = -1$ and containing $\varphi_3(2 + ai) = -1 + \frac{1}{a}i$. Namely $C_{-1}$ is the circle of radius $\frac{1}{2a}$ centered at $-1 + \frac{1}{2a}i$, and $\varphi_3(L_a) = C_{-1} - \{-1\}$.

The map $\varphi_1$ also sends $-1$ to $0$. It consequently sends the circle $C_{-1}$ to a circle $C_0$ tangent to the real line at $\varphi_1(-1) = 0$, and passing through $\varphi_1 \circ \varphi_3(3 + ai) = \frac{1}{a}i$ (since $\varphi_3(3 + ai) \in \varphi_3(L_a) \subset C_{-1}$). Namely $C_0$ is the circle of radius $\frac{1}{2a}$ centered at $\frac{1}{2a}i$.

Finally, $\varphi_3$ sends $C_1$ to a circle $C'_0$ tangent to the real line at $\varphi_3(1) = 0$, and passing through $\varphi_3 \circ \varphi_1(-3 + ai) = \frac{1}{a}i$. It follows that this circle $C'_0 = \varphi_3(C_1)$ is exactly equal to the circle $C_0 = \varphi_1(C_{-1})$.

Let $U_\infty$ be the set of points of $X$ that are on or above the line $L_a$, and let $U_0$, $U_1$ and $U_{-1}$ consist of the points of $X$ that are on or inside the circles $C_0$, $C_1$ and $C_{-1}$, respectively. In addition, because $a > 1$, the $U_i$ are disjoint. See Figure 6.19.

What the above discussion shows is that, when $P$ belongs to some of these $U_i$, and is glued to some $P' \in X$, then $P'$ must be in some other $U_j$. See Figure 6.19.

Let $U$ denote the union $U_\infty \cup U_0 \cup U_1 \cup U_{-1}$ in $X$, and let $\bar{U}$ be its image in the quotient $\bar{X}$. Namely, $\bar{U}$ is obtained from $U$ by gluing its eight sides though the restrictions of the maps $\varphi_1$, $\varphi_2 = \varphi_1^{-1}$, $\varphi_3$ and $\varphi_4 = \varphi_3^{-1}$.

![Figure 6.19](image)

Consider on $U$ the metric $d_U$ defined by the property that $d_U(P,Q)$ is the infimum of the hyperbolic length of all curves joining $P$ to $Q$ in $U$. In particular, $d_U(P,Q) = \infty$ when $P$ and $Q$ are in different $U_i$, since they cannot be joined by any curve that is completely contained in $U$.

Each $U_i$ is convex. This is fairly clear for the vertical half-strip $U_\infty$, since every hyperbolic geodesic is a circle arc centered on the $x$–axis. The property is slightly less obvious for the other $U_i$ but follows from the fact, proved below, that $U_i$ is isometric to a vertical half-strip $V_i$. A consequence of this convexity is that $d_U(P,Q) = d_{hyp}(P,Q)$ when $P$ and $Q$ are in the same $U_i$.

The metric $d_U$ induces a quotient semi-metric $\bar{d}_U$ on the quotient space $\bar{U}$.

We begin by showing that the metric space $(\bar{U}, \bar{d}_U)$ is isometric to a space that we have already encountered. Let $S_a$ be the surface of revolution obtained by revolving about the $x$–axis the portion of the pseudosphere parametrized by

$$t \mapsto (t - \tanh t, \sech t), \quad \text{arccosh} \frac{a\pi}{3} \leq t < \infty.$$
Equivalently, $S_a$ consists of those points $(x,y,z)$ of the pseudosphere $S$ such that $x \geq \log \left( \frac{a \pi}{3} + \sqrt{\frac{a^2 \pi^2}{9} - 1} \right) - \sqrt{1 - \frac{9}{a^2 \pi^2}}$.

As in Proposition 6.4, endow $S_a$ with the metric $d_{S_a}$ defined by the property that $d_{S_a}(P,Q)$ is equal to the infimum of the Euclidean lengths of all curves joining $P$ to $Q$ in $S_a$.

**Proposition 6.9.** The metric space $(U, \tilde{d}_V)$ is isometric to the space $(S_a, d_{S_a})$.

**Proof.** The proof is very similar to that of Proposition 6.4, except that we now have to glue four vertical half-strips instead of a single one.

In the hyperbolic plane $\mathbb{H}^2$, consider the subsets $V_\infty = U_\infty$, $V_{-1} = \varphi_2(U_{-1}) = \varphi_1^{-1}(U_{-1})$, $V_1 = \varphi_4(U_1) = \varphi_3^{-1}(U_1)$ and $V_0 = \varphi_4 \circ \varphi_2(U_0) = \varphi_3^{-1} \circ \varphi_1^{-1}(U_0)$.

Remember that $V_\infty = U_\infty$ is the vertical half-strip delimited by the horizontal line $L_a = \{z; \text{Im}(z) = a\}$ and by the vertical lines of equation $\text{Re}(z) = \pm 1$.

We already observed that $\varphi_1$ sends the line $L_a$ to the circle $C_1$. It also sends the points $\infty,-2$ and $-1$ to $1$, $\infty$ and $0$, respectively. Since $U_1$ is delimited by $C_1$ and by the geodesics joining $1$ to $\infty$ and $1$ to $0$, it follows that $V_1 = \varphi_1^{-1}(U_1)$ is delimited by $L_a$ and by the two geodesics joining $-1$ to $\infty$ and $-2$ to $\infty$. These hyperbolic geodesics are also the vertical lines of equations $\text{Re}(z) = -2$ and $\text{Re}(z) = -1$, so that $V_1$ is a vertical half-strip.

Similarly, since $\varphi_2(L_a) = C_{-1}$, $V_{-1} = \varphi_2^{-1}(U_{-1})$ is the vertical half-strip delimited by $L_a$ and by the two vertical lines of equations $\text{Re}(z) = 1$ and $\text{Re}(z) = 2$.

Finally, because $C_0 = \varphi_1(C_{-1}) = \varphi_1 \circ \varphi_3(L_a)$, the same type of arguments show that $V_0 = \varphi_3^{-1} \circ \varphi_1^{-1}(U_0)$ is delimited by $L_a$ and by the vertical half-lines of equations $\text{Re}(z) = 2$ and $\text{Re}(z) = 4$.

This situation is illustrated in Figure 6.19.

Let $V$ be the vertical half-strip union of $V_1, V_\infty, V_{-1}$ and $V_0$. Let $\tilde{V}$ be the quotient space obtained from $V$ by gluing its left-hand side $\{z \in V; \text{Re}(z) = -2\}$ to its right-hand side $\{z \in V; \text{Re}(z) = 4\}$ by the horizontal translation $\varphi: z \mapsto z + 6$. Since $V$ is convex, endow it with the restriction $d_V$ of the hyperbolic metric $d_{\text{hyp}}$, and let $\tilde{d}_V$ be the quotient metric induced by $d_V$ on $\tilde{V}$.

We now split the argument in two steps.

**Lemma 6.10.** The quotient space $(\tilde{V}, \tilde{d}_V)$ is isometric to the subset $(S_a, d_{S_a})$ of the pseudosphere.

**Proof.** This is essentially Proposition 6.4.
Indeed, let $V' = \{ z \in \mathbb{H}^2; 0 \leq \text{Re}(z) \leq 1, \text{Im}(z) \geq \frac{\pi}{6} \}$. Let $\tilde{V}'$ be the quotient space obtained from $V'$ by gluing its left-hand side $\{ z \in V; \text{Re}(z) = 0 \}$ to its right-hand side $\{ z \in V; \text{Re}(z) = 1 \}$ by the horizontal translation $\varphi' : z \mapsto z + 1$. Endow $V'$ with the restriction $d_{V'}$ of the hyperbolic metric $d_{\text{hyp}}$, and endow $\tilde{V}'$ with the induced quotient metric $\tilde{d}_{V'}$.

The hyperbolic isometry $\psi : z \mapsto \frac{1}{6}z + \frac{1}{3}$ sends $V$ to $V'$, and sends the gluing map $\varphi$ of $V$ to the gluing map $\varphi'$ of $V'$. Consequently $\tilde{\psi}$ induces an isometry $\tilde{\psi}$ from $(\tilde{V}, \tilde{d}_{V})$ to $(\tilde{V}', \tilde{d}_{V'})$.

Note that $V'$ is a subset of the half-strip $X_1^+$ considered in Proposition 6.4, so that $\tilde{V} \subset \tilde{X}$. The surface $S_a$ is exactly the image of $\tilde{V}'$ under the isometry $\tilde{\rho}$ constructed in the proof of Proposition 6.4. Therefore, $\tilde{\rho}$ restricts to an isometry from $(\tilde{V}', \tilde{d}_{V'})$ to $(S_a, d_{S_a})$.

The composition $\tilde{\rho} \circ \tilde{\psi}$ consequently provides an isometry from $(\tilde{V}, \tilde{d}_{V})$ to $(S_a, d_{S_a})$. \qed

**Lemma 6.11.** The quotient spaces $(\tilde{U}, \tilde{d}_{U})$ and $(\tilde{V}, \tilde{d}_{V})$ are isometric.

**Proof.** Let $\psi : U \to V$ be the map defined by the property that $\psi$ coincides with $\varphi_2 = \varphi_1^{-1}$ on $U_1$, with $\varphi_4 = \varphi_3^{-1}$ on $U_{-1}$, with $\varphi_4 \circ \varphi_2 = \varphi_3^{-1} \circ \varphi_1^{-1}$ on $U_0$, and with the identity map on $U_{\infty}$.

A case-by-case inspection of the 8 sides of $U$ that are glued together shows that $P, Q \in U$ are glued together in $\tilde{U}$ if and only if $\psi(P)$ and $\psi(Q)$ are glued together in $\tilde{V}$. For instance, if $P$ is in the intersection of $U_1$ with the vertical line of equation $\text{Re}(z) = 1$, it is glued to the point $Q = \varphi_3(P)$ contained in the intersection of $U_0$ with the geodesic joining $0$ to $-1$. Then $\psi(P) = \varphi_2(P)$ and $\psi(Q) = \varphi_4 \circ \varphi_2(Q)$ differ by the map $\varphi_4 \circ \varphi_2 \circ \varphi_3 \circ \varphi_1^{-1}$. A computation shows that $\varphi_4 \circ \varphi_2 \circ \varphi_3 \circ \varphi_2^{-1}(z) = z + 6$, so that $\psi(P)$ and $\psi(Q)$ are indeed glued together in $\tilde{V}$. The other cases are similar, and actually easier since no gluing in $\tilde{V}$ is needed.

It follows that $\psi$ induces a map $\tilde{\psi} : \tilde{U} \to \tilde{V}$, defined by the property that $\tilde{\psi}(P)$ is equal to the point of $\tilde{V}$ corresponding to $\psi(P) \in V$, for an arbitrary point $P \in U$ corresponding to $\tilde{P} \in \tilde{U}$. Let us show that $\tilde{\psi}$ is an isometry.

If $P = P_1, Q_1 \sim P_2, \ldots, Q_{n-1} \sim P_n$, $Q_n = Q$ is a discrete walk $w$ from $\tilde{P}$ to $\tilde{Q}$ in $\tilde{U}$, then $\psi(P) = \psi(P_1), \psi(Q_1) \sim \psi(P_2), \ldots, \psi(Q_{n-1}) \sim \psi(P_n), \psi(Q_n) = \psi(Q)$ is a discrete walk from $\tilde{\psi}(P)$ to $\tilde{\psi}(Q)$ in $\tilde{V}$ which has the same length as $w$. It follows that $\tilde{d}_V(\tilde{\psi}(P), \tilde{\psi}(Q)) \leq \tilde{d}_U(P, Q)$ for every $P, Q \in \tilde{U}$.

Conversely, let $\psi(P) = P'_1, Q'_1 \sim P'_2, \ldots, Q'_{n-1} \sim P'_n$, $Q'_n = \psi(Q)$ be a discrete walk $w'$ from $\tilde{\psi}(P)$ to $\tilde{\psi}(Q)$ in $\tilde{V}$. Consider the decomposition of $V$ into the four vertical half-strips $\psi(U_1), \psi(U_{\infty}), \psi(U_{-1}), \psi(U_0)$. If $P'_i$ and $Q'_i$ are not in the same half-strip $\psi(U_j)$, draw the geodesic $g$ joining $P'_i$ to $Q'_i$, consider the points $R_1, \ldots, R_k$ (with $k \leq 3$) where $g$ meets the vertical lines $\text{Re}(z) = -1, 1, 2$ separating these half-strips, and replace the part $Q'_{i-1} \sim P'_i, Q'_i \sim P'_{i+1}$ of the walk $w'$ by $Q'_{i-1} \sim P'_i, R_1 = R_1, \ldots, R_k = R_k, Q'_i \sim P'_{i+1}$. By performing finitely many such modifications we can arrange, without changing the length of $w'$, that any two consecutive $P'_i, Q'_i$ belong to the same half-strip $\psi(U_j)$. As a consequence, there exists $P_i, Q_i \in U_j$ such that $\psi(P_i) = P'_i, \psi(Q_i) = Q'_i$, and $d_U(P_i, Q_i) = d_{\text{hyp}}(P_i, Q_i) = d_{\text{hyp}}(P'_i, Q'_i)$. Then $P = P_1, Q_1 \sim P_2, \ldots, Q_{n-1} \sim P_n, Q_n = Q$ is a discrete walk from $\tilde{P}$ to $\tilde{Q}$ in $\tilde{U}$, whose length is equal to the length of $w'$. This proves that $\tilde{d}_U(P, Q) \leq \tilde{d}_V(\tilde{\psi}(P), \tilde{\psi}(Q))$ for every $P, Q \in \tilde{U}$.

This completes the proof that $\tilde{\psi} : (\tilde{U}, \tilde{d}_U) \to (\tilde{V}, \tilde{d}_V)$ is an isometry. \qed
corresponding to $P = ai$ and $Q = \frac{1}{a}i$ are very close with respect to the metric $\tilde{d}_X$, since $\tilde{d}_X(\bar{P}, \bar{Q}) \leq d_X(P, Q) = \log a^2$, but are quite far from each other with respect to the metric $\tilde{d}_U$ since there clearly is a ball $B_{d_{hyp}}(P, \varepsilon)$ which contains no point that is glued to another one, so that any discrete walk from $\bar{P}$ to $\bar{Q}$ in $\bar{U}$ has length $\geq \varepsilon$ in $\bar{U}$.

**Lemma 6.12.** If $a$ is chosen large enough that $a \log a > \frac{3}{2}$, the metrics $\tilde{d}_X$ and $\tilde{d}_U$ coincide on $\bar{U}$.

**Proof.** Because every discrete walk valued in $U$ is also valued in $X$, and because $d_X(P_i, Q_i) \leq d_U(P_i, Q_i)$ for every $P_i, Q_i \in U$, it is immediate that $\tilde{d}_U(\bar{P}, \bar{Q}) \leq \tilde{d}_X(\bar{P}, \bar{Q})$ for every $\bar{P}, \bar{Q} \in \bar{U}$.

To prove the reverse inequality, pick another number $a' > 1$ sufficiently close to 1 that $\log a' < \log a - \frac{3}{2a}$. (In particular, $a' < \frac{3}{2\log a}$.) Let $U'$ and $\bar{U}'$ be two geodesics such that, whenever the hyperbolic geodesic $\gamma_i$ joining $P_i$ to $Q_i$ meets one of the circles $C_0$, $C_1$, $C_{-1}$ and $L_a$ delimiting $U$ in $X$, it meets only at its end points $P_i, Q_i$; indeed, if this property does not hold, we can just add to the discrete walk $w$ the intersection points of $\gamma_i$ with these circles, which will not change the $d_X$-length of $w$. Adding a few more points if necessary, we can arrange that the same property holds for the circles $C_0, C_0', C_{-1}'$ and $L_{a'}$ similarly associated to $a'$.

Let $\bar{P}, \bar{Q} \in \bar{U}'$, and let $P = P_1, Q_1 \sim P_2, \ldots, Q_{n-1} \sim P_n, Q_n = Q$ be a discrete walk $w$ from $\bar{P}$ to $\bar{Q}$ in $\bar{X}$, whose $d_X$ length is such that $\ell_{d_X}(w) \leq \tilde{d}_X(\bar{P}, \bar{Q}) + \varepsilon$. Without loss of generality we can assume that, whenever the hyperbolic geodesic $\gamma_i$ joining $P_i$ to $Q_i$ meets one of the circles $C_0, C_1, C_{-1}$ and $L_a$ delimiting $U$ in $X$, it does so only at its end points $P_i, Q_i$; indeed, if this property does not hold, we can just add to the discrete walk $w$ the intersection points of $\gamma_i$ with these circles, which will not change the $d_X$-length of $w$. Adding a few more points if necessary, we can arrange that the same property holds for the circles $C_0, C_0', C_{-1}'$ and $L_{a'}$ similarly associated to $a'$.

We will show that, because of our choice of $a'$ and $\varepsilon$, the discrete walk $w$ stays in $U'$. Namely, all the $P_i$ and $Q_i$, as well as the geodesic arcs $\gamma_i$ joining them, are contained in $U'$. Indeed, suppose that the property does not hold. Let $i_1$ be the smallest index for which $\gamma_{i_1} \not\subset U$ and let $i_2$ be the largest index for which $\gamma_{i_2} \not\subset U$. By the condition that we imposed on the geodesics $\gamma_{i_1}$, the points $P_{i_1}$ and $Q_{i_1}$ are both contained in the union of the circles $C_0, C_1, C_{-1}$ and $L_a$. Similarly, let $i_1'$ be the smallest index for which $\gamma_{i_1}' \not\subset U'$ and let $i_2'$ be the largest index for which $\gamma_{i_2}' \not\subset U'$, so that $P_{i_1}'$ and $Q_{i_2}'$ belong to the union of the circles $C_0', C_1', C_{-1}'$ and $L_{a'}$. Note that $i_1 < i_1' < i_2' < i_2$.

In the quotient space $U'$, the geodesics $\gamma_i$ with $i_1 < i \leq i_1'$ project to a continuous curve $\gamma$ joining $\bar{P}_{i_1}$ to $\bar{Q}_{i_1}'$ in $U'$. Consider the isometry $\varphi: (U', d_{U'}) \to (S_{a'}, d_{S_{a'}})$ provided by Proposition 6.9. Then $\varphi(\gamma)$ is a piecewise differentiable curve joining $\varphi(\bar{P}_{i_1})$ to $\varphi(\bar{Q}_{i_1}')$ in $S_{a'}$. Note that, because $P_{i_1}$ and $Q_{i_1}'$ are in the boundary of $U$ and $U'$, respectively, $\varphi(\bar{P}_{i_1})$ and $\varphi(\bar{Q}_{i_1}')$ respectively belong to the circles $\partial S_a$ and $\partial S_{a'}$ delimiting $S_a$ and $S_{a'}$ in the pseudosphere $S$. We then use an estimate which already appeared in the proof of Lemma 6.7. Parametrize the curve $\varphi(\gamma)$ by

$$s \mapsto (t(s) - \tanh t(s), \sech t(s) \cos \theta(s), \sech t(s) \sin \theta(s)), \quad 0 \leq s \leq 1$$

for some functions $s \mapsto t(s)$ and $s \mapsto \theta(s)$, so that $\varphi(\bar{P}_{i_1}) = \partial S_a$ and $\varphi(\bar{Q}_{i_1}') = \partial S_{a'}$ respectively correspond to $s = 1$ and $s = 0$. In particular, $t(0) = \text{arccosh} \frac{a'}{3}$ and $t(1) = \text{arccosh} \frac{a}{3}$.
by definition of $S_a$ and $S_{a'}$. Then,

$$
\sum_{i=i_1}^{i_1'} d_X(P_i, Q_i) = \sum_{i=i_1}^{i_1'} \ell_{hyp}(\gamma_i)
= \ell_{euc}(\gamma) = \int_0^1 \sqrt{x'(s)^2 + y'(s)^2 + z'(s)^2} \, ds
= \int_0^1 \sqrt{t'(s)^2 \tanh^2 t(s) + \theta'(s)^2 \sech^2 t(s)} \, ds
\geq \int_0^1 t'(s) \tanh t(s) \, ds
= \log \cosh t(1) - \log \cosh t(0) = \log \frac{a}{a'}.
$$

Similarly, $\sum_{i=i_2}^{i_2'} d_X(P_i, Q_i) \geq \log \frac{a}{a'}$. As a consequence,

$$
\bar{d}_X(\bar{P}, \bar{Q}) \geq \ell_{d_X}(w) - \varepsilon = \sum_{i=1}^n d_X(P_i, Q_i) - \varepsilon
\geq \sum_{i=1}^{i_1-1} d_X(P_i, Q_i) + 2 \log \frac{a}{a'} + \sum_{i=i_2+1}^n d_X(P_i, Q_i) - \varepsilon.
$$

On the other hand, the boundary circle $\partial S_a$ has length $\frac{6}{a}$. It is consequently possible to join the two points $\varphi(\bar{P}_{i_1})$ and $\varphi(\bar{Q}_{i_2})$ by a curve of euclidean length $\leq \frac{3}{a}$ in $S_a$, so that

$$
\bar{d}_{U'}(\bar{P}_{i_1}, \bar{Q}_{i_2}) = \bar{d}_{S_a}(\varphi(\bar{P}_{i_1}), \varphi(\bar{Q}_{i_2})) \leq \frac{3}{a}.
$$

It follows that there is a discrete walk $P_{i_1} = P'_1, Q'_1 \sim P'_2, \ldots, P'_{n-1} \sim P'_{n'} = Q_{i_2}'$ of length $\leq \frac{3}{a} + \varepsilon$. Chaining this discrete walk with the beginning and the end of $w$, we obtain a discrete walk $P = P_1, Q_1 \sim P_2, \ldots, Q_{i_1-1} \sim P_{i_1}, P_{i_1} = P'_1, Q'_1 \sim P'_2, \ldots, P'_{n-1} \sim P'_{n'}, Q'_{n'} = Q_{i_2}, Q_{i_2} \sim P_{i_2+1}, \ldots, Q_{n-1} \sim P_n, Q_n = Q$ from $\bar{P}$ to $\bar{Q}$ in $\bar{X}$. Therefore,

$$
\bar{d}_X(\bar{P}, \bar{Q}) \leq \sum_{i=1}^{i_1-1} d_X(P_i, Q_i) + \frac{3}{a} + \varepsilon + \sum_{i=i_2+1}^n d_X(P_i, Q_i)
$$

Combining this with our earlier estimate for $\bar{d}_X(\bar{P}, \bar{Q})$, we conclude that $2 \log \frac{a}{a'} - \varepsilon \leq \frac{3}{a} + \varepsilon$, which is impossible by our choices of $a$, $a'$ and $\varepsilon$.

This contradiction shows that our initial hypothesis was false. Namely, the geodesics $\gamma_i$ are all in $U'$. In particular, the $P_i$ and $Q_i$ form a discrete walk from $\bar{P}$ to $\bar{Q}$ in $\bar{U}'$, so that

$$
\bar{d}_{U'}(\bar{P}, \bar{Q}) \leq \sum_{i=1}^n d_{U'}(P_i, Q_i) = \sum_{i=1}^n \ell_{hyp}(\gamma_i) = \ell_{d_X}(w) \leq \bar{d}_X(\bar{P}, \bar{Q}) + \varepsilon.
$$

Since this holds for every $\varepsilon$ that is small enough, we conclude that $\bar{d}_{U'}(\bar{P}, \bar{Q}) \leq \bar{d}_X(\bar{P}, \bar{Q})$.

Finally, the two metrics $\bar{d}_U$ and $\bar{d}_{U'}$ coincide on $\bar{U}$. This can be seen by applying the proof of Lemma 6.8 to the space $(\bar{V}, \bar{d}_V)$ of Lemma 6.11, and to $(\bar{V}', \bar{d}_{V'})$ similarly associated
to \( a' \). One can also show that the metrics \( d_{S_a} \) and \( d_{S_{a'}} \) coincide on \( S_a \) by a simple estimate of euclidean lengths of curves.

Therefore, \( \bar{d}_U(\bar{P}, \bar{Q}) \leq \bar{d}_U(\bar{P}, \bar{Q}) \leq \bar{d}_X(\bar{P}, \bar{Q}) \) for every \( \bar{P}, \bar{Q} \in \bar{U} \). Since we had already proved the reverse inequality, this shows that the metrics \( \bar{d}_U \) and \( \bar{d}_X \) coincide on \( \bar{U} \). \( \square \)

In Lemma 6.12, the condition that \( a \log a > \frac{3}{2} \) is not quite sharp. With a better hyperbolic distance estimate to improve the inequality \( \bar{d}_U(\bar{P}_i, \bar{Q}_j) \leq \frac{3}{a} \), one can show that the conclusions of Lemma 6.12 still hold for \( a \geq 2 \). This second estimate is sharp, in the sense that Lemma 6.12 fails for \( a < 2 \).

### 6.6. Triangular pillowcases

We conclude with an example where the angle condition of Theorems 5.4, 5.10 or 5.13 fails, so that we obtain surfaces with cone singularities, as in Exercise 5.8.

**Proposition 6.13.** Let \( \alpha, \beta \) and \( \gamma \) be three numbers in the interval \((0, \pi)\). Then:

1. if \( \alpha + \beta + \gamma = \pi \), there exists a triangle \( T \) of area 1 in the euclidean plane \((\mathbb{R}^2, d_{\text{euc}})\) whose angles are equal to \( \alpha, \beta \) and \( \gamma \);
2. if \( \alpha + \beta + \gamma < \pi \), there exists a triangle \( T \) in the hyperbolic plane \((\mathbb{H}^2, d_{\text{hyp}})\) whose angles are equal to \( \alpha, \beta \) and \( \gamma \);
3. if \( \pi < \alpha + \beta + \gamma < \pi + 2 \min\{\alpha, \beta, \gamma\} \), there exists a triangle \( T \) in the sphere \((\mathbb{S}^2, d_{\text{sph}})\) whose angles are equal to \( \alpha, \beta \) and \( \gamma \).

In addition, in each case, the triangle \( T \) is unique up to isometry of \((\mathbb{R}^2, d_{\text{euc}}), (\mathbb{H}^2, d_{\text{hyp}})\) of \((\mathbb{S}^2, d_{\text{sph}})\), respectively.

**Proof.** The euclidean case is well-known. See Exercises 6.15 and 4.5 for a proof in the hyperbolic and spherical cases. \( \square \)

See also Exercises 3.15 and 4.6 for a proof that the conditions of Proposition 6.13 are necessary.

Given \( \alpha, \beta \) and \( \gamma \in (0, \pi) \), let \( T \) be the euclidean, hyperbolic or spherical triangle provided by Proposition 6.13.

Choose an isometry \( \varphi \) of \((\mathbb{R}^2, d_{\text{euc}}), (\mathbb{H}^2, d_{\text{hyp}})\) or \((\mathbb{S}^2, d_{\text{sph}})\), according to the case, such that \( \varphi(T) \) is disjoint from \( T \). The existence of \( \varphi \) is immediate in the euclidean and hyperbolic space. For the spherical case, note from the proof of Proposition 6.13(3) in Exercise 4.5 that \( T \) is always contained in the interior of a hemisphere, so that we can use for \( \varphi \) the reflection across the great circle delimiting this hemisphere.

Let \( X \) be the union of \( T \) and \( \varphi(T) \). We can consider \( X \) as a non-connected polygon, whose edges are the three sides \( E_1, E_3, E_5 \) of \( T \) and the corresponding three sides \( E_2 = \varphi(E_1), E_4 = \varphi(E_3) \) and \( E_6 = \varphi(E_5) \) of \( \varphi(T) \). We can then glue these edges by the gluing maps \( \varphi_1: E_1 \to E_2, \varphi_3: E_3 \to E_4, \varphi_5: E_5 \to E_6 \) defined by the restrictions of \( \varphi \) to the edges indicated.

By Theorems 5.3, 5.9 or 5.12, this gluing data provides a quotient metric space \((\bar{X}, \bar{d}_X)\). This metric space is locally isometric to \((\mathbb{R}^2, d_{\text{euc}}), (\mathbb{H}^2, d_{\text{hyp}})\) or \((\mathbb{S}^2, d_{\text{sph}})\) everywhere, except at the three points that are the images of the vertices under the quotient map. The metric has cone singularities at these three points, with respective cone angles \( 2\alpha, 2\beta, 2\gamma < 2\pi \), in the sense of Exercise 5.8.

Note that \( \bar{X} \) is obtained by gluing two triangles \( T \) and \( \varphi(T) \) along their edges. In the euclidean case, this is the familiar construction of a pillow case obtained by sewing together
two triangular pieces of cloth. Figure 6.21 attempts to describe the geometry of $\bar{X}$ in all cases.

![Figure 6.21. Triangular pillowcases](image)

**Exercises for Chapter 6**

**Exercise 6.1.** We want to rigorously prove that the metric spaces $(\bar{X}_1, \bar{d}_{X_1})$ and $(\bar{X}_2, \bar{d}_{X_2})$ of Section 6.1.1 are homeomorphic. Without loss of generality, we can assume that the lower left corners of the rectangle $X_1$ and the parallelogram $X_2$ are equal to the origin $(0,0) \in \mathbb{R}^2$.

a. Show that there is a unique linear map $\mathbb{R}^2 \to \mathbb{R}^2$ that sends the bottom edge of $X_1$ to the bottom edge of $X_2$, and the left edge of $X_1$ to the left edge of $X_2$. Show that this linear map restricts to a homeomorphism $\varphi: \bar{X}_1 \to \bar{X}_2$.

b. Show that two points $P$ and $Q \in X_1$ are glued together to form a single point of $\bar{X}_1$ if and only if their images $\varphi(P)$ and $\varphi(Q)$ are glued together in $X_2$. Conclude that $\varphi: \bar{X}_1 \to \bar{X}_2$ induces a map $\bar{\varphi}: \bar{X}_1 \to \bar{X}_2$, defined by the property that $\bar{\varphi}(P) = \varphi(P)$ for every $P \in X_1$.

c. Show that $\bar{\varphi}: \bar{X}_1 \to \bar{X}_2$ is a homeomorphism.

**Exercise 6.2.** Let $X$ be the torus obtained by gluing opposite sides of a euclidean parallelogram $X$. Let $\varphi: X \to X$ be the rotation of angle $\pi$ around the center of the parallelogram, namely around the point where the two diagonals meet. Show that there exists a unique isometry $\bar{\varphi}: \bar{X} \to \bar{X}$ of the quotient metric space $(\bar{X}, \bar{d}_X)$ such that $\bar{\varphi}(\bar{P}) = \bar{\varphi}(P)$ for every $P \in \bar{X}$.

**Exercise 6.3.** Let $(\bar{X}, \bar{d}_X)$ be the torus obtained by gluing opposite edges of a euclidean rectangle $X = [a,b] \times [c,d] \subset \mathbb{R}^2$ by translations, where $X$ is endowed with the euclidean metric $d_X = d_{euc}$. Recall that a geodesic of $(\bar{X}, \bar{d}_X)$ is a curve $\gamma$ in $\bar{X}$ such that, for every $P \in \gamma$ and every $Q \in \gamma$ sufficiently close to $P$, there is a piece of $\gamma$ joining $P$ to $Q$. Show that every closed geodesic curve in $\bar{X}$ has length $\geq 1$. Possible hint: Consider the projection of the pre-image $P^{-1}(\gamma) \subset X_1 \subset \mathbb{R}^2$ for each of the coordinate axes.

c. Consider the case where $X$ is the square $X_1 = [0,1] \times [0,1]$. Show that every closed geodesic curve in $X_1$ has length $\geq 1$. Possible hint: Consider the projection of the pre-image $P^{-1}(\gamma) \subset X_1 \subset \mathbb{R}^2$ for each of the coordinate axes.

c. Consider the case where $X$ is the square $X_1 = [0,\frac{1}{2}] \times [0,2]$. Show that $(\bar{X}, \bar{d}_X)$ contains a closed geodesic of length $\frac{1}{2}$.

d. Conclude that the euclidean tori $(\bar{X}_1, \bar{d}_{X_1})$ and $(\bar{X}_2, \bar{d}_{X_2})$ are not isometric.

**Exercise 6.4.** Let $(\bar{X}, \bar{d}_X)$ be the torus obtained by gluing together opposite sides of the square $X = [0,1] \times [0,1]$. Let $\pi: X \to \bar{X}$ denote the quotient map.

a. Given $a \in [0,1]$, let $\varphi_a: X \to X$ be the (discontinuous) function defined by the property that $\varphi_a(x,y) = (x+a, y)$ if $0 \leq x \leq 1-a$ and $\varphi_a(x,y) = (x+a-1, y)$ if $1-a < x \leq 1$. Show that $\varphi_a$ induces a map $\pi \circ \varphi_a: \bar{X} \to \bar{X}$, uniquely determined by the property that $\varphi_a \circ \pi = \pi \circ \varphi_a$.

b. Show that, if $w$ is a discrete walk $P = P_1, Q_1 \sim P_2, \ldots, Q_{n-1} \sim P_n, Q_n = Q$ is a discrete walk from $P$ to $Q$ in $\bar{X}$, there exists a discrete walk $w'$ from $\varphi_a(P)$ to $\varphi_a(Q)$ which has the same $\bar{d}_X$-length $\ell_{\bar{d}_X}(w') = \ell_{\bar{d}_X}(w)$ as $w$. Hint: When $P_i$ and $Q_i$, are on opposite sides of the line $x = 1-a$, add to $w$ the point where the line segment $[P_i,Q_i]$ meets this line.

c. Use Part b to show that $\bar{d}_X(\varphi_a(P), \varphi_a(Q)) \leq \bar{d}_X(P, Q)$ for every $P, Q \in \bar{X}$. Then show that $\varphi$ is an isometry of $(\bar{X}, \bar{d}_X)$.

d. Use the above construction to show that the metric space $(\bar{X}, \bar{d}_X)$ is homogeneous namely that, for every $P, Q \in \bar{X}$, there exists an isometry of $(\bar{X}, \bar{d}_X)$ which sends $P$ to $Q$. 

\[ \alpha + \beta + \gamma = \pi \quad \alpha + \beta + \gamma < \pi \quad \alpha + \beta + \gamma > \pi \]
e. More generally, let \( Y \) be a parallellogram in \( \mathbb{R}^2 \) and let \((\bar{Y}, d_\bar{Y})\) be the quotient metric space obtained by gluing opposite sides of \( Y \) by translations as in Section 6.1.1. Show that \((\bar{Y}, d_\bar{Y})\) is homogeneous.

**Exercise 6.5.** Let \((\bar{X}, d_\bar{X})\) be the Klein bottle obtained by gluing together the sides of the square \( X = [0, 1] \times [0, 1] \) as in Section 6.1. Let \( \pi : X \to \bar{X} \) denote the quotient map.

a. Recall that a geodesic of \((\bar{X}, d_\bar{X})\) is a curve \( \gamma \) in \( \bar{X} \) such that, for every \( P \in \gamma \) and every \( \bar{Q} \in \gamma \) sufficiently close to \( P \), there is a piece of \( \gamma \) joining \( P \) to \( \bar{Q} \) which is the shortest curve going from \( P \) to \( \bar{Q} \), where the length of a curve in \((\bar{X}, d_\bar{X})\) is defined as in Exercise 2.11. Show that the image \( \gamma_1 = \pi([0, 1] \times \{0\}) \) is a closed geodesic of length 1, and that there exists at least one closed geodesic of length 2 which is disjoint from \( \gamma_1 \). Show that \( \gamma_2 = \pi([0, 1] \times \{\frac{1}{2}\}) \) satisfies the same two properties, namely is a closed geodesic of length 1 which is disjoint from a closed geodesic of length 2.

b. Show that \( \gamma_1 \) and \( \gamma_2 \) are the only two closed geodesics in \( \bar{X} \) satisfying the properties of Part a. Hint: For a closed geodesic \( \gamma \) in \( \bar{X} \), note as in Exercise 6.3, that the preimage \( \pi^{-1}(\gamma) \) must consist of parallel line segments in the square \( X \).

c. Conclude that the Klein bottle \( \bar{X} \) is not homogeneous.

**Exercise 6.6.** In the euclidean plane, let \( X_1 \) be the square with vertices \( (0, 0), (0, 1), (1, 0), (1, 1) \); let \( X_2 \) be the parallellogram with vertices \( (0, 0), (0, 1), (1, 0), (1, 1) \); and let \( X_3 \) be the parallellogram with vertices \( (0, 0), (0, 1), (1, 2), (2, 3) \). Let \( \bar{X}_1, \bar{X}_2 \) and \( \bar{X}_3 \) be the euclidean tori obtained by gluing opposite sides of these parallelograms. Show that the euclidean tori \( \bar{X}_1, \bar{X}_2, \bar{X}_3 \) are all isometric. Hint: Show that each of these euclidean tori is obtained by gluing the sides of two suitably chosen triangles. (You may need to use the result of Exercise 5.4 to justify the fact that the order of the gluings does not matter.)

**Exercise 6.7.** Given a euclidean hexagon \( X \), index its vertices as \( P_1, P_2, \ldots, P_6 \) in this order as one goes around the hexagon. Suppose that opposite edges have the same length, and that the sum of the angles of \( X \) at its odd vertices \( P_1, P_3 \) and \( P_5 \) is equal to \( 2\pi \). Use elementary euclidean geometry to show that opposite edges of \( X \) are parallel.

**Exercise 6.8.** A surface is **non-orientable** if it contains a subset homeomorphic to the Möbius strip. For instance, the Klein bottle is a non-orientable euclidean surface. Construct a non-orientable hyperbolic surface.

**Exercise 6.9.** The surface of genus \( g \) is the immediate generalization of the case \( g = 2 \), and consists of a torus with \( g \) handles. Construct a hyperbolic surface of genus \( g \) for \( g = 3, 4 \), and then for any \( g \geq 2 \). Possible hint: Glue opposite sides of a hyperbolic polygon with suitably chosen angles; Proposition 6.13 may be convenient to construct this polygon.

**Exercise 6.10 (The projective plane).** Let \( X \) be the spherical polygon of Section 6.3, with the gluing data indicated there, and let \((\bar{X}, d_\bar{X})\) be the corresponding quotient space. Let \( \mathbb{R}P^2 \) be the set of lines passing through the origin \( O \) in \( \mathbb{R}^3 \). For any two \( L, L' \in \mathbb{R}P^2 \), let \( \theta(L, L') \in [0, \frac{\pi}{2}] \) be the angle between these two lines in \( \mathbb{R}^3 \).

a. Show that \( \theta \) defines a metric on \( \mathbb{R}P^2 \).

b. Consider the map \( \varphi : X \to \mathbb{R}P^2 \) which to \( P \in X \) associates the line \( OP \). Show that \( \varphi \) induces a bijection \( \bar{\varphi} : \bar{X} \to \mathbb{R}P^1 \).

c. Show that this bijection \( \bar{\varphi} \) is an isometry between the metric spaces \((\bar{X}, d_\bar{X})\) and \((\mathbb{R}P^2, \theta)\).

d. Show that the metric space \((\mathbb{R}P^2, \theta)\) (and consequently \((\bar{X}, d_\bar{X})\) as well) is homogeneous.

**Exercise 6.11.** Let \((\bar{X}, d_\bar{X})\) be the projective plane of Section 6.3 (or Exercise 6.10 above). Show that, for every great circle \( C \) of \( \mathbb{S}^2 \), the image of \( X \cap C \) under the quotient map \( \pi : X \to \bar{X} \) is a closed geodesic of length \( \pi \). Conclude that the spherical surface \((\bar{X}, d_\bar{X})\) is not isometric to the sphere \((\mathbb{S}^2, d_{\mathbb{S}^2})\).

**Exercise 6.12.** Let \((\bar{X}, d_\bar{X})\) be the projective plane constructed in Section 6.3, by gluing onto itself the boundary of a hemisphere \( X \subset \mathbb{S}^2 \). Let \( P_0 \) be the center of this hemisphere, namely the unique point such that \( X = B_{\mathbb{S}^2}(P_0, \frac{\pi}{2}) \). Show that \( \bar{X} - P_0 \) is homeomorphic to the Möbius strip of Section 6.4.

**Exercise 6.13.** Show that the “hyperbolic square” of Figure 6.18 really has the symmetries of a square, in the sense that there exists an isometry of \( \mathbb{H}^2 \) which sends 0 to 1, to \( \infty \), to \(-1 \) and \(-1 \) to 0.

**Exercise 6.14.** Let \( X \) be the hyperbolic polygon of Figure 6.18, namely the infinite square in \( \mathbb{H}^2 \) with vertices at infinity 0, 1, \( \infty \), and \(-1 \). We will glue its edges in a different way from the construction of Section 6.5. Namely, we glue the edge \( \infty \) to the edge \( 0 \) by the map \( z \mapsto \frac{1}{-z+2} \), and the edge \(-1 \) to the edge \( 0 \) by \( z \mapsto \frac{1}{-z-2} \).

Let \((\bar{X}, d_\bar{X})\) be the quotient metric space provided by this gluing construction.

a. Show that \((\bar{X}, d_\bar{X})\) is locally isometric to \((\mathbb{H}^2, d_{\mathbb{H}^2})\).

b. Show that \( \bar{X} \) can be decomposed as \( \bar{X} = \bar{X}_0 \cup \bar{U}_1 \cup \bar{U}_2 \cup \bar{U}_3 \) where each \((\bar{U}_i, d_{\bar{X}})\) is isometric to a subset \((S_a, d_{S_a})\) of the pseudosphere as in Proposition 6.9, and where \( \bar{X}_0 \) is the image of a bounded subset \( X_0 \subset X \) under the quotient map \( X \to \bar{X} \). Hint: Adapt the analysis of Section 6.5.
c. Show that \((\bar{X}, d_{\bar{X}})\) is homeomorphic to the complement of three points in the sphere \(S^2\).

**EXERCISE 6.15 (Hyperbolic triangles).** Let \(\alpha, \beta, \gamma\) be three positive numbers with \(\alpha + \beta + \gamma < \pi\). We want to show that there exists a hyperbolic triangle in \(\mathbb{H}^2\) whose angles are equal to \(\alpha, \beta, \gamma\), and that this triangle is unique up to isometry of \(\mathbb{H}^2\). For this, we just adapt the proof of Lemma 6.2.

Let \(g\) be the complete hyperbolic geodesic of \(\mathbb{H}^2\) with end points \(0\) and \(\infty\). Let \(h\) be the complete hyperbolic geodesic passing through the point \(i\) and such that the angle from \(g\) to \(h\) at \(i\), measured counterclockwise, is equal to \(+\beta\). For \(y < 1\), let \(k_y\) be the complete geodesic passing through \(iy\) and such that the counterclockwise angle from \(g\) to \(k_y\) at \(iy\) is equal to \(-\gamma\). Compare Figure 6.9.

a. Show that the set of those \(y < 1\) for which \(k_y\) meets \(h\) is an interval \((y_0, 1)\).

b. When \(y\) is in the above interval \((y_0, 1)\), let \(\alpha_y\) be the counterclockwise angle from \(h\) to \(k_y\) at their intersection point. Show that

\[
\lim_{y \to -1^-} \alpha_y = \pi - \beta - \gamma \quad \text{and} \quad \lim_{y \to y_0^+} \alpha_y = 0.
\]

Conclude that there exists a value \(y \in (y_0, 1)\) for which \(\alpha_y = \alpha\).

c. Show that the above \(y \in (y_0, 1)\) with \(\alpha_y = \alpha\) is unique.

d. Let \(T\) and \(T'\) be two hyperbolic triangles with the same angles \(\alpha, \beta, \gamma\). Show that there is an isometry \(\varphi\) of the hyperbolic plane \((\mathbb{H}^2, d_{\text{hyp}})\) such that \(\varphi(T) = T'\). Hint: For \(g\) and \(h\) as above, apply suitable hyperbolic isometries to send \(T\) and \(T'\) to triangles with one edge in \(g\) and another edge in \(h\), and use Part c.

**EXERCISE 6.16 (The Gauss-Bonnet formula).** Let \(X\) be a bounded polygon in \(\mathbb{R}^2\), \(\mathbb{H}^2\) or \(S^2\) consisting of finitely many disjoint convex polygons \(X_1, X_2, \ldots, X_n\). Let \(\bar{X}\) be obtained by gluing together pairs of edges of \(X\). As usual we assume that, for every vertex \(P\) of \(X\), the angles of \(X\) at the vertices that are glued to \(P\) add up to \(2\pi\), so that \(\bar{X}\) is a euclidean, hyperbolic or spherical surface. The quotient space \(\bar{X}\) is now decomposed into \(m\) images of the convex polygons \(X_1, n\) images of the \(2n\) edges of \(X\), and \(p\) points images of the vertices of \(X\). The **Euler characteristic** of \(\bar{X}\) is the integer

\[
\chi(\bar{X}) = m - n + p.
\]

A deep result, which we cannot prove here, asserts that the Euler characteristic \(\chi(\bar{X})\) is independent of the way the space \(\bar{X}\) is obtained by gluing edges of a polygon in the sense that, if two such surfaces \(\bar{X}\) and \(\bar{X}'\) are homeomorphic, they have the same Euler characteristic \(\chi(\bar{X}) = \chi(\bar{X}')\). See for instance [Massey, Chap. IX, §4] or [Hatcher, §2.2]

a. Compute the Euler characteristic of the torus, the Klein bottle, the surface of genus 2, and the projective plane.

b. Show that, for each of the convex polygons \(X_i\),

\[
\sum_{j=1}^{n_i} \theta_j = (n_i - 2)\pi + K \text{Area}(X_i),
\]

where: \(X_i\) has \(n_i\) vertices, and \(\theta_1, \theta_2, \ldots, \theta_{n_i}\) are the angles of \(X_i\) at these vertices; if \(X_i \subset \mathbb{R}^2\) is a euclidean polygon, \(\text{Area}(X_i)\) denotes its usual area and \(K = 0\); if \(X_i \subset \mathbb{H}^2\) is a hyperbolic polygon, \(\text{Area}(X_i)\) denotes its hyperbolic area \(\text{Area}_{\text{hyp}}(X_i)\) as defined in Exercise 3.14, and \(K = -1\); if \(X_i \subset S^2\) is a spherical polygon, \(\text{Area}(X_i)\) denotes its surface area in \(S^2\) and \(K = +1\). Hint: Use the convexity of \(X_i\) to decompose it into \((n_i - 2)\) triangles, and apply the results of Exercises 3.15 and 4.6 to these triangles.

c. Use Part b to show that

\[
2\pi \chi(\bar{X}) = K \text{Area}(X).
\]

In particular, if a surface \(\bar{X}\) is obtained by gluing together the sides of a euclidean, hyperbolic, or spherical polygon \(X\) as above, its Euler characteristic \(\chi(\bar{X})\) must be zero, negative or positive, respectively. The equation of Part c is a special case of a more general statement known as the **Gauss-Bonnet formula**.