1.1. An imaginary world

HENRI POINCARÉ did not invent non-Euclidean geometry. Even the famous Poincaré disk existed before him. However, his use of this geometry and its group of isometries was so staggering that the name Poincaré disk is by no means undeserved. The goal of this chapter is not to describe the history of non-Euclidean geometry, because a number of excellent works are already devoted to that topic (see [16, 23, 29, 47, 48, 51], among my favorites). I would have liked to propose a guided tour of the disk, but it is much too vast and I myself have only explored a small part. Instead, I invite the reader on an excursion. If one takes a random walk in the Euclidean plane, the risk is that one will return to the point of origin over and over ceaselessly, but this risk does not exist in non-Euclidean geometry! We shall see that a random path in the disk does not take many side trips and almost surely leads somewhere. My main goal is to try to convey a geometric intuition for this object which has progressively passed, in less than two centuries, from the status of a counterexample—whose very existence was doubtful—to that of a central concept that has invaded nearly all of mathematics.

Of course, this text is not directed to the experts. I have instead tried to make it accessible to a (motivated) undergraduate. For those who would like to know more, I provide a copious bibliography—in reality, a pretext to present some of my favorite books.

Let us first make the acquaintance of a non-Euclidean geometry by using the imagery that Poincaré created in Science and Hypothesis [62, chap. 3]:

"THE NON-EUCLIDEAN WORLD

Let us assume [...] a world enclosed in [a large circle] and subject to the following laws:

The temperature in this world is not uniform; it is largest at the center, and it diminishes as one moves away from the center, so that it reduces to absolute zero when one reaches [the circle] where this world is enclosed.

I will moreover specify the following law by which this temperature varies. Let \( R \) be the radius of the limit [circle]; let \( r \) be the distance from the point under consideration to the centre of [this circle]. The absolute temperature will be proportional to \( R^2 - r^2 \).

I will additionally assume that, in this world, all bodies have the same coefficient of dilatation, in such a way that the length of any ruler shall be proportional to its absolute temperature."
Finally, I will assume that an object transported from one point to another, whose temperature is different, shall immediately reach thermal equilibrium with its new location.

Nothing in these hypotheses is contradictory or unimaginable.

A moving object will then become smaller and smaller as it approaches the limit [circle].

Let us first observe that, if this world is finite from the point of view of our customary geometry, it will appear infinite to its inhabitants.

In fact, when they wish to approach the limit [circle], they will get colder and become smaller and smaller. The steps they take are therefore also smaller and smaller, so that they can never reach the limit [circle].

If, for us, geometry is merely the study of the laws by which non-deformable solids may move; for these imaginary beings, it will be the study of the laws that drive solids deformed by these differences in temperature about which I have just spoken. [...] For brevity, I shall, with the reader’s permission, call such a movement a non-Euclidean displacement.

Thus beings like us, whose education would take place in such a world, would not have the same geometry as us. [If these imaginary beings] founded a geometry, [...] it would be non-Euclidean geometry."

It is in this world “neither contradictory nor unimaginable” that we shall take our excursion.

Here is a quote from Coxeter showing just how “real” this geometry is for mathematicians [23]:

When Hamlet exclaims (in Act II, Scene II) “I could be bounded in a nutshell and count myself a king of infinite space” he is providing a poetic anticipation of Poincaré’s inversive model of the infinite hyperbolic plane, using a circular “nutshell” for the Absolute.

The reader is therefore warned that this world is vast...  

1.2. Some formulas

Before beginning our excursion, we must introduce some definitions, notations, and formulas which will serve to paraphrase the imagery of Poincaré’s description.

We denote by $D = \{ z \in \mathbb{C} \mid |z| < 1 \}$ the open unit disk in the complex plane. If $v$ is a tangent vector to $D$ at a point $z$, with Euclidean norm $\|v\|_{\text{Eucl}}$, then its hyperbolic norm $\|v\|_{\text{hyp}}$ is defined by

$$\|v\|_{\text{hyp}} = \frac{1}{1 - |z|^2} \|v\|_{\text{Eucl}}.$$  

---

It appears that Klein is responsible for the term “hyperbolic geometry”. There are of course good reasons for this choice, but one can only regret the all-too-frequent use of the word “hyperbolic” in mathematics, with quite different meanings.
From this we define the **hyperbolic length** of a curve \( \gamma : [0, 1] \rightarrow \mathbb{D} \) by
\[
\text{length}_{\text{hyp}}(\gamma) = \int_0^1 \left\| \frac{d\gamma}{dt} \right\|_{\text{hyp}} dt,
\]
and the **hyperbolic distance** (or **Poincaré distance**) \( \text{dist}_{\text{hyp}}(z_0, z_1) \) between two points \( z_0 \) and \( z_1 \) of the disk as the minimum of the hyperbolic lengths of curves joining \( z_0 \) and \( z_1 \). The **Poincaré disk** is the metric space thus obtained.

Why choose this factor \( 1/(1 - |z|^2) \) rather than another? Quite simply because it is the “obvious” factor that guarantees that the object we have just defined is **homogeneous**. If \( \alpha \) is a real number and \( a \) is an element of \( \mathbb{D} \), the transformation of \( \mathbb{C} \cup \{\infty\} \) defined by
\[
f_{\alpha, a}(z) = \exp(i\alpha) \frac{z - a}{1 - \overline{a}z}
\]
preserves the unit disk (check this!). The set of these transformations forms a group that we will often encounter in this chapter. Let us observe for the moment that this group acts **transitively** on the disk: given any two points in \( \mathbb{D} \), one of the group elements sends the first point to the second point (check this!). The derivative of \( f_{\alpha, a} \) is
\[
f'_{\alpha, a} = \exp(i\alpha) \frac{1 - |a|^2}{(1 - \overline{a}z)^2},
\]
from which it follows that
\[
\frac{|df_{\alpha, a}(z)|}{1 - |f_{\alpha, a}(z)|^2} = \frac{|dz|}{1 - |z|^2}
\]
(again, check this...). In other words, the **Poincaré metric** is invariant under the group of the \( f_{\alpha, a} \)'s, and the disk is seen to be homogeneous: *its group of isometries acts transitively* and all of its points are equivalent.

We will see a bit later that this homogeneity is a crucial property that “almost” characterizes the Poincaré disk. For the moment, we will use the homogeneity in order to effortlessly obtain certain formulas that we will need.

It is easy to find the shortest curve (in the hyperbolic sense) joining the point 0 of \( \mathbb{D} \) to the point \( r \) situated on the axis \([0, 1] \subset \mathbb{D} \). In fact, it suffices to observe that if \( \gamma : [0, 1] \rightarrow \mathbb{D} \) joins 0 to \( r \), the radial projection \( |\gamma| : [0, 1] \rightarrow [0, 1] \subset \mathbb{D} \) also joins 0 to \( r \), and that its hyperbolic length is less than that of \( \gamma \) (because the radial component of a vector is shorter than the vector itself). Thus, the unique curve minimizing the hyperbolic length between 0 and \( r \) is the radius \([0, r]\) with hyperbolic length
\[
\int_0^r \frac{dt}{1 - t^2} = \frac{1}{2} \log \left( \frac{1 + r}{1 - r} \right) = \tanh^{-1}(r).
\]
This length tends to infinity as \( r \) tends to 1, as Poincaré’s imagery describes.

To find the **geodesics** of \( \mathbb{D} \), i.e., the curves with fixed endpoints that minimize the hyperbolic lengths, it suffices to apply the group of isometries. Given two points \( z_0 \) and \( z_1 \), we can find an isometry that sends the first to 0 and the second to a point of the positive real axis, as previously. The sought-after geodesic is therefore the image of a radial segment by an isometry; it is an arc of a circle orthogonal to the unit circle (or a radial segment). To be convinced of this, the reader should
Remember (or check for himself) that a projective transformation
\[ z \in \mathbb{C} \cup \{\infty\} \mapsto \frac{Az + B}{Cz + D} \in \mathbb{C} \cup \{\infty\} \]
(with \( A, B, C, \) and \( D \) complex numbers such that \( AD - BC \neq 0 \)) sends a circle to a circle (in the Riemann sphere \( \mathbb{C} \cup \{\infty\} \)) and preserves the angle of intersection between two circles. Thus, a diameter of the unit disk is sent by an isometry \( f_{a,a} \) onto an arc of a circle orthogonal to the unit circle (or onto another diameter). In the same way, it is easy to use homogeneity to establish the formula that gives the distance between two points \( z_0 \) and \( z_1 \). We already know that
\[ \text{dist}_{\text{hyp}}(0, r) = \frac{1}{2} \log \left( \frac{1 + r}{1 - r} \right) = \tanh^{-1}(r) \]
which can also be written
\[ \frac{1}{2} \left| \log \left( \frac{0 - (-1)}{0 - 1} : \frac{r - 1}{r - (-1)} \right) \right| \]
where we recognize \([-1 : 1 : 0 : r] \), the cross ratio of four points in \( \mathbb{C} \cup \{\infty\} \). Recall that the cross ratio of four distinct points \( x, y, z, t \) is defined by
\[ [x : y : z : t] = \frac{z - x}{z - y} : \frac{t - y}{t - x} \]
and that for any projective transformation \( f \), we have
\[ [f(x) : f(y) : f(z) : f(t)] = [x : y : z : t]. \]
Since the projective transformations \( f_{a,a} \) preserve the cross ratio, hyperbolic distances, and circle orthogonal to the unit circle, we easily obtain the hyperbolic distance between any two points \( z_0 \) and \( z_1 \) in \( \mathbb{D} \). It suffices to consider the unique circle (or diameter) orthogonal to the unit disk that contains \( z_0 \) and \( z_1 \). This circle meets the unit circle in two points \( u, v \), and we have
\[ \text{dist}_{\text{hyp}}(z_0, z_1) = \frac{1}{2} \left| \log[u : v : z_0 : z_1] \right|. \]
The reader who prefers a formula that only brings \( z_0 \) and \( z_1 \) into play may express (as Poincaré did in his original article on the subject (61)) the distance as a function of the cross ratio of the (cocyclic) points \( z_0, 1/\overline{z_0}, z_1, \) and \( 1/\overline{z_1} \).
Let us observe that the disk is in fact 2-homogeneous: if \( \text{dist}_{\text{hyp}}(z_0, z_1) = \text{dist}_{\text{hyp}}(z'_0, z'_1) \), there exists an isometry that sends \( z_0 \) to \( z'_0 \) and \( z_1 \) to \( z'_1 \).

### 1.3. The Poincaré disk is ubiquitous

We know (at least since Euclid!) the geometry of the Euclidean plane. The geometry of the sphere is equally familiar because, after all, *geo-metry* is the science that *measures* the *earth*. It took two millennia for hyperbolic geometry to install herself among mathematicians (and a little among physicists). However, she deserves the same respect as her two older sisters.

It seems to me that the following “characterization theorem” fully justifies this respect. Its statement is simple, but its proof is not. It uses some of the most difficult theorems of the 20th century and goes well beyond the level of this elementary chapter. In the appendix, I will nevertheless try to give a few indications of this proof.

**Theorem.** Let \( (X, d) \) be a metric space with the following properties:

- \( (X, d) \) is a *surface*: every point of \( X \) has a neighborhood homeomorphic to an open subset of \( \mathbb{R}^2 \);
- \( (X, d) \) is 2-*homogeneous*: if \( d(x, y) = d(x', y') \), there exists an isometry that sends \( x \) to \( x' \) and \( y \) to \( y' \);
- \( (X, d) \) is a *geodesic space*: for any pair of points \( (x, y) \), there exists a curve \( \gamma : [0, \ell] \to X \) such that \( \gamma(0) = x, \gamma(1) = y \), and which is an isometry onto its image (so that \( d(x, y) = \ell \)).

Then \( (X, d) \) is isometric to one of the following three examples:

- the *Euclidean plane*;
- the *sphere* of radius \( R \) in Euclidean space, or the quotient of this sphere in which one identifies pairs of diametrically opposed points (the *Klein elliptic space*);
- the *Poincaré disk* endowed with a constant multiple of hyperbolic distance.

A few comments on the hypotheses of this theorem.

The property of having dimension 2 is fundamental. Later we will describe other geodesic, 2-homogeneous metric spaces with larger dimension. The fact of being a surface, on the other hand, is not very important: one could assume, for example, that \( X \) is locally compact and has topological dimension \(^2\ 2 \).

Homogeneity is of course essential, but 2-homogeneity is much less so. We will see that if we replace the second hypothesis by simple homogeneity, the theorem remains true as long as we add a few additional, less important examples (can the reader can what they are?).

Nor is the third hypothesis very important. Let us remark that if \( (X, d) \) is a metric space, then \( (X, \phi(d)) \) is also a metric space, inducing the same topology, just as homogeneous as the first, provided that \( \phi : \mathbb{R}^+ \to \mathbb{R}^+ \) is subadditive, that is, such that \( \phi(s + t) \leq \phi(s) + \phi(t) \). If we suppress the third hypothesis, then the theorem’s conclusion is hardly weakened at all: the space \( (X, d) \) is then obtained by this construction from one of the preceding examples: Euclidean, elliptic, or hyperbolic.

\(^2\)A topological space has topological dimension less than or equal to \( d \) if it admits arbitrarily fine coverings such that the intersection of \( (d + 2) \) distinct open sets in a given covering is empty. For example, the product of two graphs has dimension 2 but is not, in general, a surface.
We shall come back to these “details” further on, but let us retain the fundamental fact that the Euclidean, elliptic, and hyperbolic geometries are essentially the only homogeneous metric surfaces.

In truth, these three geometries $(X, d)$ are much more than homogeneous. If $Y$ is any piece of $X$ and if $f : Y \to X$ is an isometry onto its image, then $f$ is the restriction of a global isometry of $X$. If $Y$ contains only one point, this is the usual property of homogeneity; if $Y$ contains two points, this is 2-homogeneity, etc.

1.4. Lots and lots of models

Homogeneous objects of dimension 2 arise naturally in many situations. The preceding theorem then allows us to identify such an object with one of the three geometries. Thus we find models of hyperbolic geometry abounding in the literature. Even though the goal of this chapter is not to discuss the epistemological aspect of non-Euclidean geometry, let us remark upon the common usage of the word “model” in this context: it is as though this geometry had an “ideal” intrinsic existence that mathematicians seek to understand by constructing several “illustrations” of it.

I will describe here a few of these models, but there are plenty of others (see for example [2, 47, 69, 72]).

The first is nothing but a benign change of variables: the transformation

$$ z \mapsto i \frac{1 + z}{1 - z} $$

sends the disk $D$ onto the Poincaré half-plane $\mathbb{H} = \{ z \in \mathbb{C} | \operatorname{Im} z > 0 \}$ because

$$ \operatorname{Im} \left( i \frac{1 + z}{1 - z} \right) = \frac{1 - |z|^2}{|1 - z|^2}. $$

In this half-plane, the hyperbolic metric becomes $|dz|/\operatorname{Im} z$, and the geodesics are half-circles (or rays) orthogonal to the boundary (see Fig. 2).

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure2}
\caption{Triangle in the half-plane}
\end{figure}

---

3\footnote{One sometimes speaks of the Lobachevskii plane, which is of course entirely justified, but in a work devoted to the heritage of Poincaré...}

4\footnote{The reader who works through the calculations will find that a factor of 2 (or of 1/2) appears when passing between this metric and the one on p. 18. This is slightly unfortunate, albeit interesting, but for purposes of this exposition can be ignored. –trans.]
This model has the advantage of presenting the group of orientation-preserving isometries in the simplest way:

$$z \in \mathbb{H} \mapsto \frac{Az + B}{Cz + D} \in \mathbb{H}$$

with $A$, $B$, $C$, and $D$ real numbers such that $AD - BC = 1$. We recognize the group $\text{PSL}(2, \mathbb{R})$.

With the disk and half-plane models, one can already carry out many calculations. One may easily convince oneself that Euclid’s fifth postulate is not true in this geometry: given a point $x$ on the exterior of a geodesic $\gamma$, there are infinitely many geodesics that contain $x$ and are “parallel” to $\gamma$, that is, they do not intersect $\gamma$ (FIG. 3).

**Figure 3.** The fifth postulate

_Here is a less well-known model._ We consider the set $\mathcal{E}$ of ellipses in the Euclidean plane, centered at the origin, and bounding a region of unit area. We may think of such an ellipse $e$ as the unit sphere of a norm $\| \cdot \|_e$ on $\mathbb{R}^2$. If $e_1$ and $e_2$ are two elements of $\mathcal{E}$, we may compare them by setting

$$\text{dist}(e_1, e_2) = \log \sup_{v \in \mathbb{R}^2 \setminus \{(0,0)\}} \frac{\|v\|_{e_1}}{\|v\|_{e_2}}.$$

With just a moment of reflection, the reader can verify that:

- $\text{dist}$ is indeed a (geodesic?) distance;
- $(\mathcal{E}, \text{dist})$ is a surface (because two parameters suffice to describe an element of $\mathcal{E}$);
- $(\mathcal{E}, \text{dist})$ is homogeneous (there is only one ellipse, up to affine transformations).

Another moment of reflection will equally convince the reader that $(\mathcal{E}, \text{dist})$ is neither Euclidean nor elliptic. Hence it must be a model of the Poincaré disk!

It remains to exhibit an explicit isometry between $(\mathcal{E}, \text{dist})$ and $(\mathbb{D}, \text{dist}_\text{hyp})$. This, of course, poses no problem. To each point of the disk having the form $\rho \exp(i\phi)$, we associate the ellipse of ellipticity $^5 \rho < 1$ whose major axis makes an angle of $\phi/2$ with the real axis. The verification that this bijection is indeed an isometry is a routine exercise.

This model suggests a generalization: the space of symmetric convex sets having volume 1 in $\mathbb{R}^n$ is naturally a metric space.

---

$^5$The ellipticity of an ellipse with axes $b \leq a$ is $(a - b)/(a + b)$. Do not confuse it with the eccentricity.
One other model: the Hilbert metric [44, App. 1]. Let $C$ be a bounded open convex set in $\mathbb{R}^n$. If $x$ and $y$ are two points of $C$, the line $(xy)$ that joins them intersects $C$ on an open interval $(u, v)$ (see Fig. 4). We set

$$\text{dist}_{\text{Hilbert}}(x, y) = |\log[u : v : x : y]|$$

where this time we use the cross ratio of the four points $u, v, x, y$ on the real line (before, we were using the cross ratio of four points on the complex line).

![Figure 4. Hilbert distance](image)

Here again, it is not difficult to assure oneself that $\text{dist}_{\text{Hilbert}}$ is a geodesic distance (contemplate Fig. 4, not forgetting that the cross ratio is invariant under projection). Evidently, this distance is invariant under the group of projective transformations that preserve $C$. In the particular case where $C$ is an ellipse, this group acts transitively on $C$ (exercise)$^6$ and the metric space we obtain is homogeneous. We have thus found a version of the Poincaré disk in the interior of an ellipse, often called the “Klein model”. It is not so easy to find an isometry between $(C, \text{dist}_{\text{Hilbert}})$ and $(\mathbb{D}, \text{dist}_{\text{hyp}})$, however. Note that the two models are very different: the geodesics in the Klein model are line segments, while those in the Poincaré disk are arcs of circles. It is not a priori clear that there is a homeomorphism of the disk that turns line segments into arcs of circles. Here is one method: given real numbers $a$ and $b$, consider the second degree equation in an unknown $z$ of the form $(1 - a)z^2 + 2bz + (1 + a) = 0$, with discriminant $4(a^2 + b^2 - 1)$. Thus to each point $(a + ib)$ in the unit disk, we can associate a second degree equation that has a unique solution $z_{a+ib}$ in the upper half-plane $\mathbb{H}$. It is a pleasant exercise to show that the transformation $(a + ib) \in \mathbb{D} \mapsto z_{a+ib} \in \mathbb{H}$ realizes an isometry between the Klein metric and the Poincaré metric (as seen in the half-plane).

The Hilbert metric has many other interesting properties (see [43] for numerous examples and developments). Note that the ellipse is not the only projectively

---

$^6$We cite here the lovely consequence Hilbert drew from this result: it is impossible to construct the center of a circle using only a ruler. In fact, if such a construction existed, its conjugate by a projective transformation preserving the circle could be used to construct any point of the disk: a contradiction!
homogeneous convex figure: so is the interior of the triangle\(^7\). What is the Hilbert metric in this case?

### 1.5. Attempts to visualize in space

Even though Riemann taught us that one can (should?) do geometry on an abstract Riemannian manifold, many of us like to visualize surfaces as embedded in Euclidean space.

A differentiable function \(\phi : \mathbb{D} \to \mathbb{R}^N\) is called an \textit{isometric immersion} if for any vector \(v\) tangent to the disk, we have \(\|v\|_{\text{hyp}} = \|d\phi(v)\|_{\text{Eucl}}\). If, in addition, \(\phi\) is injective, we call it an \textit{isometric embedding}. This does not mean that \(\phi\) is an isometry onto its image, but only that the hyperbolic length of a curve in \(\mathbb{D}\) equals the Euclidean length of the image of the curve by \(\phi\). Moreover, there is no function \(\phi : \mathbb{D} \to \mathbb{R}^N\) that is an isometry onto its image, because such a function would have to send a geodesic onto a line, and the image of \(\mathbb{D}\) would therefore have to be contained in a Euclidean plane, which is of course impossible.

Early on, Beltrami sought to construct an isometric embedding of the hyperbolic disk into \(\mathbb{R}^3\). By a very simple method, he succeeded in \textit{locally} finding such an embedding. We consider a surface of revolution whose equation in cylindrical coordinates \((r, z)\) has the form \(r = F(z)\). In order to find \(F\) such that the surface is isometric to the Poincaré disk, we must solve a second-order differential equation. Among its solutions, we find the \textit{tractrix}, the curve followed by an object pulled by a rope of constant length and whose free end moves along a line (FIG. 5). Thus the

![Figure 5. Tractrix](image)

\textit{tractrix} of revolution, often called a \textit{pseudosphere} (FIG. 6), is a local model of the Poincaré disk [57].

![Figure 6. Pseudosphere](image)

\(^7\)Thus the foregoing argument of Hilbert shows that the centroid of a triangle cannot be constructed using only a ruler; this is perhaps more surprising than for the center of a circle?
This tractrix contains a singular point (a cusp), however, in such a way that the surface of revolution is also singular along a cusp circle. Therefore we only obtain a piece of a surface isometric to a piece of the disk. The left part of FIG. 7 (taken from [65]) shows a disk in $\mathbb{D}$ with a small radius and its isometric image in the pseudosphere. The right part shows a part of $\mathbb{D}$ (with area $2\pi$) which is isometric to the complement of a generatrix in a half-pseudosphere.

![Figure 7. Pseudosphere and disk](image)

Moreover, we cannot escape these singularities: Hilbert proved\(^8\) that there is no isometric embedding of the the Poincaré disk into 3-dimensional Euclidean space, such that the embedding is of class $C^2$. The proof is quite clever: it consists of a detailed analysis of two families of asymptotic lines\(^9\) on a surface in $\mathbb{R}^3$ which is locally isometric to the disk. These curves form a Tchebychev net, which was a notion introduced for a very concrete problem in the article entitled “Sur la coupe des vêtements” (“The cutting out of clothes”) [70]. One considers a piece of fabric in the plane formed by the interlaced threads

$$x = i/N, \quad y = j/N, \quad i, j = 0, \ldots, N \quad (N \text{ large}).$$

Next one deforms the fabric in space in such a way that the sides of each stitch, initially in the shape of a square, keep the same length. In other words, one considers the surfaces “clothed” in the fabric $u : [0, a] \times [0, b] \to \mathbb{R}^3$ such that $\partial u/\partial x$ and $\partial u/\partial y$ have norm 1 everywhere (but are not necessarily orthogonal). One finds that the asymptotic lines of a surface locally isometric to the Poincaré disk clothe this surface in the preceding sense (see FIG. 8).

This is the point at which Hilbert’s proof begins. For the (interesting) continuation, see [20, 29, 44]. It is worth remarking that this problem of Tchebychev continues to develop: apart from the fact that it is intimately linked to the so-called sine-Gordon partial differential equations, one uses this kind of ideas in the construction of certain contemporary materials (see, for example, [63]).

The proof of Hilbert’s theorem could not be too elementary, because the theorem is false for class $C^1$. It follows from a theorem of Nash that there is a $C^1$ isometric embedding of the Poincaré disk into the space $\mathbb{R}^3$ [36, 53].

What do these embeddings look like? It is very useful to construct concrete models of them, for example out of paper, as Beltrami did in the 19th century (FIG. 9).

---

\(^8\)In truth, Hilbert assumed the embedding to be analytic, and it was only much later that the $C^2$ version was proved.

\(^9\)Let $m$ be a point on a smooth surface $S$ embedded in $\mathbb{R}^3$. Denote by $\Pi$ the tangent plane to $S$ at $m$ and by $N$ the normal line. Locally, $S$ is the graph of a function $u$ from a neighborhood of $m$ in $\Pi$ to $N$. The second derivative of $u$ at $m$ is the second fundamental form at the point $m$. The isotropic directions of this quadratic form, if they exist, are the asymptotic directions, and the curves that are everywhere tangent to one of these directions are the asymptotic lines of $S$. 
Take a large number of equilateral cardboard triangles, say a few centimeters to the side, and tape their edges together in such a way that each vertex is surrounded by 7 triangles: you will get an approximation of a piece of the Poincaré disk (FIG. 10). Notice that if you replace 7 with 6, you will end up constructing a Euclidean plane with the usual tiling by equilateral triangles. By using 5 triangles around each vertex, one obtains a sphere, tiled like an icosahedron. (What happens with 4, 3, 2 triangles, with 8, 9, etc.?)

The remarkable internet site [69] provides a visual introduction to hyperbolic geometry. Another site [71] contains a virtual exhibit that seeks to develop one’s intuition for this type of objects. There is even a set of instructions for crocheting a Poincaré disk!
In [60], the authors do not hesitate to evoke a “jupe à godets” [“gored skirt”] (FIG. 12) and compare the pseudosphere to a lily (FIG. 13).

![Figure 12. Jupe à godets](image1)

![Figure 13. Calla lily](image2)

Of course, it is possible to apply general theorems about $C^\infty$ isometric embeddings of Riemannian manifolds into Euclidean spaces of sufficiently large dimension. Following Gromov [36] (generalizing work of Nash), a $k$-dimensional Riemannian manifold has a $C^\infty$ isometric embedding into $(k+2)(k+3)/2$-dimensional Euclidean space. As for $k$-dimensional hyperbolic space (here we have only discussed the dimension 2 case), there is a $C^\infty$ isometric embedding into $(5k-5)$-dimensional Euclidean space. In fact, Blanuša constructed in 1955 an explicit $C^\infty$ embedding of the Poincaré disk into $\mathbb{R}^6$ [14]. Can the disk be isometrically immersed in $\mathbb{R}^4$? This seems to be an open question. Gromov showed, however, that any compact piece of the disk can be isometrically immersed in an arbitrary\(^{10}\) (non-empty!) open subset of $\mathbb{R}^4$.

Such embeddings are in fact of little use, because they cannot be natural. A natural embedding would be an embedding $i : \mathbb{D} \to \mathbb{R}^N$ such that for any isometry $f$ of $\mathbb{D}$, there would be an isometry $f$ of $\mathbb{R}^N$ such that $i \circ f = f \circ i$. Such an embedding cannot exist for purely algebraic reasons: there is no non-trivial homomorphism from the group of (orientation-preserving) isometries of the disk into the group of isometries of Euclidean space (exercise, not so easy!).

On the other hand, there do exist natural embeddings into infinite dimensional space. Later I will show an explicit construction of an embedding $i : \mathbb{D} \to H$, where $H$ is a Hilbert space, satisfying the following properties:

- $\text{dist}_{\text{hyp}}(x, y) = \|i(x) - i(y)\|^2$ for all $x, y \in D$;
- $i$ is natural in the above sense.

1.6. A little bit of triangle geometry

The most famous theorem of hyperbolic geometry is due to Gauss [32]:

**Theorem.** The area of a triangle with angles $\alpha, \beta, \gamma$ is $\pi - (\alpha + \beta + \gamma)$.

\(^{10}\)One might believe that the diameter of the open set would necessarily be larger than the diameter of the compact set, but nothing of the sort is true: an isometric embedding is not an isometry! For example, an isometric embedding of $\mathbb{R}$ into $\mathbb{R}^3$ is just an embedded curve, parametrized by arclength; it is not difficult to “wind up” such a curve into an arbitrarily small ball.
Before sketching two proofs of this theorem (there are plenty of others), we must explain the words “triangle, angle, area”. The Poincaré disk is a Riemannian manifold whose metric \( ds = |dz|/(1 - |z|^2) \) is conformal to the Euclidean metric \( |dz| \); this means that the angle between two tangent vectors at the same point of the disk is the same regardless of whether one calculates with \( ds \) or \( |dz| \). The hyperbolic area element is given by \( d\text{area}_{\text{hyp}} = (1 - |z|^2)^{-2} \, d\text{area}_{\text{Eucl}} \).

A triangle must of course be understood to be defined by three points and bounded by three geodesic segments; it therefore has three angles and an area.

It is easy to convince oneself that the sum of the angles of a triangle is less than \( \pi \) (put one of the vertices at the center of the disk and compare the angles of the hyperbolic triangle with those of the Euclidean triangle having the same vertices).

The first sketch of a proof of the theorem may require a few extra developments to confirm it, but it is rather intuitive. If \( P \) is a geodesic polygon with \( n \) sides and whose angles measure \( \alpha_1, \ldots, \alpha_n \), we set
\[
\mathcal{A}(P) = (n - 2)\pi - \sum_{i=1}^{n} \alpha_i.
\]

Then we remark that

\begin{itemize}
  \item if we cut \( P \) along a geodesic to obtain polygons \( P_1 \) and \( P_2 \), we evidently have \( \mathcal{A}(P) = \mathcal{A}(P_1) + \mathcal{A}(P_2) \) (see FIG. 14);
  \item if \( P \) is a triangle with small diameter, \( \mathcal{A}(P) \) is small (“because” in a small neighborhood, a Riemannian metric is “almost” Euclidean: this is a point that must be made more precise);
  \item \( \mathcal{A} \) is clearly invariant under isometries.
\end{itemize}

Thus we can mimic the classical construction of Lebesgue measure on the plane. We use \( \mathcal{A}(P) \) as the measure of a polygon, and we define the measure of a Borel set using coverings by polygons. Thus, \( \mathcal{A} \) defines a measure on the disk that is invariant by isometries. Homogeneity shows that \( \mathcal{A} \) must therefore be a constant.
multiple of the hyperbolic area. Therefore we have “established” that
\[
\text{area}(P) = c \left( (n - 2)\pi - \sum_{i=1}^{n} \alpha_i \right)
\]
and all that is left is to determine the constant \( c \). We will not do this here, partly because the calculation has no particular interest; but especially because in multiplying the Poincaré metric by a strictly positive constant \( \lambda \), the constant \( c \) is divided by \( \lambda^2 \). In fact, the choice of the constant in the definition of the Poincaré metric is mostly dictated by the desire to normalize the constant \( c \) to 1, which is possible because we know \( c \) is strictly positive.

Here is a second, more convincing proof. Three points on the boundary of \( \mathbb{D} \) determine an ideal triangle whose “vertices” are “at infinity” (see FIG. 15).

These ideal triangles, despite not being bounded, have bounded area. This follows from the fact that two geodesics converging to the same point on the boundary approach each other exponentially fast (in the hyperbolic sense, of course), so that the area integral converges. Since the group of isometries of the disk acts transitively on triplets of boundary points (exercise), all of these ideal triangles have the same area. One finds the value of this area to be \( \pi \) (calculate this!).

![Figure 15. Ideal triangle](image1)

![Figure 16. \( T(\alpha) \)](image2)

![Figure 17. \( T(\alpha + \beta) \)](image3)

![Figure 18. Ideal hexagon](image4)
Now let us consider the triangle $T(\alpha)$ with one angle $\alpha \in [0, \pi]$ and whose two other vertices are at infinity (FIG. 16). FIG. 17 shows that

$$\text{area}(T(\alpha + \beta)) = \text{area}(T(\alpha)) + \text{area}(T(\beta)) - \pi,$$

from which it follows that $F(\alpha) = \pi - \text{area}(T(\alpha))$ satisfies $F(\alpha + \beta) = F(\alpha) + F(\beta)$. Since $F$ is continuous, there exists a constant $c$ such that $F(\alpha) = c\alpha$. As $\text{area}(T(\pi)) = 0$, we obtain $c = 1$ and $\text{area}(T(\alpha)) = \pi - \alpha$.

Finally let us consider a “real” triangle $T(\alpha, \beta, \gamma)$ having all three vertices at “finite distance” (FIG. 18). By extending the sides to infinity, we obtain an ideal hexagon.

Each vertex of the triangle determines two isometric copies of $T(\alpha), T(\beta),$ and $T(\gamma)$ respectively, one of which contains $T(\alpha, \beta, \gamma)$. The area of the hexagon is therefore

$$2[(\pi - \alpha) + (\pi - \beta) + (\pi - \gamma)] - 2\text{area}(T(\alpha, \beta, \gamma)).$$

Moreover, as this hexagon can evidently be decomposed into four ideal triangles, its area is $4\pi$. As announced before, we obtain

$$\text{area}(T(\alpha, \beta, \gamma)) = \pi - (\alpha + \beta + \gamma).$$

I propose an “application exercise” to end this section. Euclid (almost) showed in the Elements that if two polygons $P$ and $Q$ in the Euclidean plane have the same area, one can cut each of them into a finite number of pieces $P_1, \ldots, P_n$ and $Q_1, \ldots, Q_n$ so that for all $i$ the “pieces” $P_i$ and $Q_i$ are isometric [31, Book VI]. Show that the same is true in the Poincaré disk. This property is no longer true in higher dimensions and leads to fascinating developments about Hilbert’s third problem: can two polyhedra with the same volume in Euclidean space be cut into isometric pieces? See [15] for an elementary introduction and [21] for more information.

1.7. The disk is a tree

It follows from Gauss’s formula that the area of a triangle is bounded by $\pi$. Strange geometry in which triangles can be arbitrarily large in size, but whose area is bounded!

The hyperbolic area of a disk of radius $\rho$ is easy to calculate:

$$A(\rho) = \int_0^{\tanh \rho} \frac{2\pi t}{(1 - t^2)^2} \, dt = \frac{\pi}{2} (\cosh \rho - 1).$$

We thus notice that this area tends to infinity when the radius tends to infinity (exponentially fast—we will return to this point). It follows from this that the radius $\rho$ of the circle inscribed in a triangle is bounded independently of the triangle, since $\frac{\pi}{2} (\cosh \rho - 1) \leq \pi$ leads to $\rho \leq 1$ (FIG. 19).

**Definition.** We say that a geodesic metric space $(X, d)$ is $\delta$-hyperbolic if, for any triple of points $(x, y, z)$ and any choice of geodesics $[x, y], [y, z], [z, x]$ pairwise connecting them, every point of $[x, y]$ is at a distance less than $\delta$ from a point of $[x, z]$ or of $[z, y]$.

Gromov extracted this geometric property from the Poincaré disk and recognized that this benign definition captures the essence of this geometry [37]. A bounded metric space $(X, d)$ is of course $\delta$-hyperbolic, with $\delta = \text{diam}(X, d)$, but this particular case evidently holds no interest. The theory is only interesting for unbounded spaces. We have seen, for example, that the Poincaré disk is $\delta$-hyperbolic.
with $\delta = 2$. Numerous metric spaces are $\delta$-hyperbolic and are therefore cousins of the disk. Regretfully, I will nonetheless remain in the disk!

Another way to express the $\delta$-hyperbolicity of the Poincaré disk is to say that the union $[x, y] \cup [y, z] \cup [z, x]$ of the three sides of a triangle with vertices $x, y, z$ is at a bounded distance from the union $[x, o] \cup [y, o] \cup [z, o]$, where $o$ denotes the center of the inscribed circle: the union of the three sides is nearly a $Y$ (see FIG. 20).

One finds that this property generalizes to any finite piece, which can be approximated by a finite tree$^{11}$, with an error that depends only on the number of points.

Let us make this statement more precise. If we choose an arbitrary length for each edge of a finite tree, we can construct a metric realization of the tree: we attach Euclidean segments having the chosen lengths by their endpoints, following the combinatorics of the tree. By definition, the distance between two points is the length of the shortest path connecting them. We call a metric space constructed in this way a metric tree (see FIG. 21).

\begin{figure}[h]
\centering
\includegraphics[width=0.8\textwidth]{metric_tree}
\caption{Metric tree}
\end{figure}

**Proposition.** We consider $n$ points $x_1, \ldots, x_n$ in a $\delta$-hyperbolic metric space $(X, d)$. Then there exists a metric tree $(T, d_T)$ and $n$ points $x'_1, \ldots, x'_n$ in $T$ such that

$$d(x_i, x_j) - 100 \delta \log n \leq d_T(x'_i, x'_j) \leq d(x_i, x_j)$$

for all $i, j$.

$^{11}$A tree is a connected graph with no cycles.
1.7. THE DISK IS A TREE

In the photograph of “hyperbolic lettuce” (FIG. 22), one plainly sees arborescent nerves that approximate the whole leaf.

![Hyperbolic Lettuce Image](image)

**FIGURE 22.** Hyperbolic lettuce

I reiterate that the “lack of isometry” $100\delta \log n$ depends (a little) on $n$ but not on the points $x_i$, which could be very far away from each other.

The proof of this property is not difficult, but it is austere. The reader may begin by trying to construct it for himself then, in case of failure, he may go consult [33, 37] to appreciate the virtuosity with which Gromov manipulates the triangle inequality!

Thus, if we want to draw figures in the Poincaré disk formed from a large number of points very distant from each other, the result is nearly a tree. Very often, this gives a good intuition of hyperbolic geometry (FIG. 23).

![Cactus Image](image)

**FIGURE 23.** Cactus approaching the geometry of the disk

If $(X, d)$ is a $\delta$-hyperbolic metric space, one can divide the distance by a (large) constant $k > 0$ and set $d_k = d/k$. This comes back to looking at the space $(X, d)$.
“from far away”, and the resulting metric space is of course \((\delta/k)\)-hyperbolic. If \(k\) approaches infinity, the defect \(100(\delta/k) \log n\) of approximation by trees tends to 0 (\(n\) being fixed), that is to say “the space \((X, d_k)\) tends to a tree.” One can give a precise meaning to the previous sentence. I will not do so, however, because that would lead us to a discussion of Hausdorff–Gromov topology on the space of metric spaces, ultrafilters, etc. The interested reader may consult [39].

Whatever the case, let us retain that the Poincaré disk, seen from far away, resembles a tree.

### 1.8. Some examples of dendrologic intuition

Dendrologic\(^{12}\) geometry is often intuitive. We shall see, using several simple examples, how it can guide our understanding of hyperbolic geometry.

**The Pythagorean theorem.** Let us compare right triangles in the Euclidean plane, in a tree, and in the disk.

When we say the triangle \(ABC\) “has a right angle at \(A\)”, we mean that \(A\) is the foot of the altitude from \(C\) to the line \((AB)\), or equivalently that \(A\) is the closest point to \(C\) and lying on \((AB)\). In dendrologic geometry, right triangles are therefore those for which the shortest path from \(C\) to \(B\) passes through \(A\), in other words those for which \(BC = AB + AC\) (fig. 24).

![Figure 24. Right triangles](image)

In trees, the Pythagorean theorem has “lost its square”: the hypotenuse is the sum of the sides. What is it for the Poincaré disk? One finds in books the hyperbolic Pythagorean theorem:

\[
\cosh BC = \cosh AB \cdot \cosh AC.
\]

If we take into account the fact that \(\cosh x \simeq \exp(x)/2\) for large real \(x\), we find that \(BC \simeq AB + AC\) as our dendrologic intuition had suggested.

**Growth and transience.** Because the Poincaré disk is homogeneous, we should compare it with a homogeneous tree. Let us consider, for example, the infinite homogeneous tree all of whose vertices have valence 3 (i.e., each is incident with three edges) (fig. 25).

The ball of radius \(n\) centered at a point \(x_0\) contains \(3 \cdot 2^{n-1}\) vertices. We have rediscovered what we met before: the volume of a ball grows exponentially as a function of its radius.

Now let us consider a *random walk* in the tree. A point starts at the vertex \(x_0\) and jumps randomly each second to one of the three neighboring vertices, with equal probability. Let us denote \(d(n)\) the distance between the point \(x_n\) at the

---

\(^{12}\)Dendrology from *dendro-* and *-logy*, n. The botanical study of trees and other woody plants. *(American Heritage Dictionary)* [The original text refers to *Le Petit Robert*. –trans.]
moment $n$ and the starting point $x_0$. If $d(n) \neq 0$, then $d(n + 1) = d(n) + 1$ with probability $2/3$ and $d(n + 1) = d(n) - 1$ with probability $1/3$. On the other hand, if $d(n) = 0$, then we have $d(n + 1) = 1$. There is therefore a definite tendency to run off to infinity. It is not difficult to conclude that almost surely:

- $d(n)/n$ tends to $1/3$ (in particular $d(n)$ tends to infinity);
- the geodesic $[x_0, x_n]$ tends to an infinite geodesic $\gamma = [x_0, x_\infty)$ and the distance between $x_n$ and $\gamma$ is bounded by a constant multiple of $\log n$.

The same thing happens in the Poincaré disk for Brownian motion. For any Riemannian manifold $M$ and for any point $x_0$, there exists a probability measure, called Wiener measure, on the space of paths $\gamma : \mathbb{R}^+ \to M$ starting from $x_0$ (see for example [11, 27]). For the Poincaré disk:

- almost every curve $\gamma : \mathbb{R}^+ \to M$ starting from a fixed point converges to a point of the boundary $\omega(\gamma) \in \partial \mathbb{D}$;
- the hyperbolic distance between $\gamma(t)$ and the geodesic $[x_0, \omega(\gamma)]$ is bounded by a constant multiple of $\log t$.

The proof is of course more difficult than in the case of trees, but the underlying idea is the same.

The fact that a random walk runs off to infinity can also be expressed in terms of the behavior of the solutions to the heat equation as time tends to infinity. D. Sullivan explained to me that it is almost impossible to heat houses in the Poincaré disk, because you cannot stop the heat from escaping to infinity!

**Quasi-geodesics.** One of the attractive features of hyperbolic geometry is its robustness. We shall illustrate it here in a simple but fundamental example.

A curve $\gamma : \mathbb{R} \to X$ in a metric space $(X, d)$ is a quasi-geodesic if there exist $a$ and $b > 0$ such that, for all $t_1, t_2$, we have:

$$a^{-1}|t_1 - t_2| - b \leq d(\gamma(t_1), \gamma(t_2)) \leq a|t_1 - t_2| + b.$$  

**THEOREM.** Every quasi-geodesic in a $\delta$-hyperbolic metric space is within a bounded distance of a geodesic.

The idea of the proof consists first of studying the case of a tree. A quasi-geodesic in a tree can cross several times over the same vertex, but these events cannot happen at very different times ($\gamma(t_1) = \gamma(t_2)$ leads to $|t_1 - t_2| \leq ab$). The
result is that a quasi-geodesic in a tree consists in fact of a geodesic onto which are grafted some round trips of bounded length (FIG. 27). The general case is nothing but an adaptation of the particular case of the tree (see for example [33]).

Here is a sample application. We consider a Riemannian metric $g$ on the disk whose ratio with the hyperbolic metric is bounded. This means that there exists a constant $C_t > 1$ such that for any tangent vector $v$, the ratio between the $g$-norm and the hyperbolic norm of $v$ is bounded by $C_t^{-1}$ and $C_t$. Then a geodesic $\gamma: \mathbb{R} \to D$ for the metric $g$ is evidently a quasi-geodesic for the hyperbolic metric: it therefore remains within a bounded distance from a geodesic of the hyperbolic metric. This permits us a canonical way of associating a Poincaré geodesic to a geodesic\footnote{In this text, a geodesic is by definition a curve that minimizes the length between any two points, even greatly distant ones. One normally speaks of minimizing geodesics.} of $g$. This is the starting point for the phenomenon of structural stability of the geodesic flow of negatively curved compact manifolds: the qualitative behavior of geodesics does not depend (too much) on the choice of metric. Surrounding this is a long history which would carry us too far from the quasi-geodesic that we are trying to
1.9. The disk is a curve

follow in this chapter, and would lead us from Hadamard to Gromov passing by Anosov [5, 35, 42]…

**Sensitivity to initial conditions.** Let us consider two geodesics \( \gamma_1, \gamma_2 : \mathbb{R}^+ \to \mathbb{R}^2 \) in the Euclidean plane (that is, two rays). Choose \( \varepsilon > 0 \) (small) and \( T > 0 \) (large). Suppose that \( \gamma_1 \) and \( \gamma_2 \) coincide at \( t = 0 \) and are close at time \( T \), that is

\[
\text{dist}_{\text{Eucl}}(\gamma_1(T), \gamma_2(T)) \leq \varepsilon.
\]

Then, by similar triangles\(^{14}\)

\[
\text{dist}_{\text{Eucl}}(\gamma_1(2T), \gamma_2(2T)) \leq 2\varepsilon.
\]

In other words, after doubling the amount of time, the distance between \( \gamma_1(t) \) and \( \gamma_2(t) \) has only doubled: it therefore remains small (FIG. 28-a).

Now let us consider the case of a tree. FIG. 28-b shows two geodesics starting from the same point that coincide up to time \( T \) and diverging afterwards, in the sense that \( \text{dist}(\gamma_1(2T), \gamma_2(2T)) = 2T \). The fact that \( \gamma_1 \) and \( \gamma_2 \) remain close (and even coincide) on \([0, T]\) does not lead to any estimate for \( \text{dist}(\gamma_1(2T), \gamma_2(2T)) \) (other than the triangle inequality). No similar triangles in dendrologic geometry!

The same phenomenon occurs in the disk. Let us consider two geodesics \( \gamma_1 \) and \( \gamma_2 : \mathbb{R}^+ \to D \) such that \( \gamma_1(0) = \gamma_2(0) \) and \( \text{dist}_{\text{hyp}}(\gamma_1(T), \gamma_2(T)) = \varepsilon \). It is possible to calculate \( F(\varepsilon, T) = \text{dist}_{\text{hyp}}(\gamma_1(T), \gamma_2(T)) \) explicitly thanks to the formulas of hyperbolic trigonometry (see [2, 23]). One finds \( F(\varepsilon, T) = 2\sinh^{-1}(2\sinh \frac{\varepsilon}{2} \cosh T) \). If \( \varepsilon \) and \( T \) are small, we recover similar triangles: \( F(\varepsilon, T) \approx 2\varepsilon \). In contrast, if \( \varepsilon > 0 \) is fixed and \( T \) tends to infinity, \( F(\varepsilon, T) \approx 2T \), as for a tree (FIG. 28-c).

This is the simplest example of sensitivity to initial conditions. If one has a precise knowledge of a geodesic on an interval \([0, T]\) at one’s disposal, it is impossible to deduce from this a precise knowledge on the interval \([T, 2T]\). The future seems to have forgotten the past. This is one of the most important ideas hiding behind the concept of deterministic chaos. The geodesics of the disk are deterministic in the sense that they are completely determined by their initial position and velocity, but their behavior is unpredictable in practice [8, 25, 34, 42, 67].

---

\(^{14}\)I have chosen throughout this section to refer to similar triangles where the original has “le théorème de Thaîs”, because the property being invoked is not commonly known by the name of Thales’ theorem in English-speaking countries, at least in the U.S. and the U.K. –trans.]
Let \( f \) be an orientation-preserving isometry of the Poincaré disk. By composing \( f \) with an appropriate \( f_{a,a} \), we may suppose that \( f \) leaves the origin fixed and that its derivative at this point is the identity. Being an isometry, \( f \) is also the identity along the geodesic starting from 0 with an arbitrary initial direction \( v \). Consequently, \( f \) is the identity everywhere, and the initial isometry was indeed of the form \( f_{a,a} \).

Let \( f \) be a holomorphic bijection of the disk. By composing \( f \) with an appropriate \( f_{a,a} \), we may suppose that \( f \) leaves the origin fixed. The classical Schwarz’s lemma then asserts that \( |f(z)| \leq |z| \) for every point \( z \) of the disk \([3, 4, 52, 66]\). By considering the inverse \( f^{-1} \), we obtain that in fact \( |f(z)| = |z| \) and that \( f \) is a rotation. The initial holomorphic bijection was indeed of the form \( f_{a,a} \).

Thus, the disk \( \mathbb{D} \) can be endowed with a(n oriented) metric structure and with a holomorphic structure whose automorphism groups coincide. It is for this reason that the relationships between hyperbolic geometry (in real dimension 2) and holomorphic geometry (in complex dimension 1) are so close.

The omnipresence of the Poincaré disk as a metric space has a holomorphic counterpart: it concerns the famous uniformization theorem, probably one of the most beautiful mathematical jewels discovered in the 19th century, the result of efforts by Gauss, Riemann, Schwarz, Klein, Koebe, and Poincaré.

A Riemann surface is a holomorphic manifold of dimension 1. In other words, it is a topological space covered by open sets (“chart domains”) homeomorphic to open sets in \( \mathbb{C} \) so that the coordinate changes are holomorphic. Riemann surfaces are “curves” because they have complex dimension 1, but they are surfaces of real dimension 2. It is this curve–surface duality that gives the theory its flavor. The literature on this subject is immense, but I particularly recommend [18, 22, 41, 49, 64], among others...

A Riemann surface is simply connected if it is connected and if every closed curve can be continuously deformed to a point.

**Theorem** (Uniformization theorem). Let \( S \) be a simply connected Riemann surface. Then \( S \) is biholomorphically equivalent to the complex plane \( \mathbb{C} \), the Riemann sphere \( \mathbb{C} \cup \{ \infty \} \), or the Poincaré disk \( \mathbb{D} \).

One must carefully distinguish this theorem from the particular case, taught in undergraduate courses, which asserts that a simply connected open set in the complex plane (non-empty and not all of \( \mathbb{C} \)) is biholomorphically equivalent to the disk. The uniformization theorem deals with an abstract Riemann surface, which cannot a priori be embedded in the plane. One may in particular apply it to the universal cover of an arbitrary Riemann surface, for example compact (that is, an algebraic curve, following Riemann). For a proof, one may consider for example [41]. This concerns one of Poincaré’s initial motivations: a compact Riemann surface with genus greater than or equal to 2 can be identified with a quotient of the disk by a Fuchsian group, a discrete group of isometries.

**Fig. 29** contains two illustrations of this kind of group, drawn by J. Leys [26, 55]; the second was inspired by the work of the artist M. C. Escher [17, 24, 30].

With regret, I will not discuss these groups, which however would deserve it. See for example [24, 56, 59, 72].

The connection between holomorphic structure and hyperbolic metric is summarized by Schwarz’s lemma as expressed intrinsically by Pick. Every holomorphic function \( f : \mathbb{D} \rightarrow \mathbb{D} \) is contracting for the hyperbolic metric: for any \((z_1, z_2)\), we
have \( \text{dist}_{\text{hyp}}(f(z_1), f(z_2)) \leq \text{dist}_{\text{hyp}}(z_1, z_2) \). In truth, Pick’s contribution was not limited to expressing Schwarz’s lemma in an invariant way. He solved the difficult problem of “holomorphic interpolation”: if \( z_1, \ldots, z_n \) and \( w_1, \ldots, w_n \) are two \( n \)-tuples of points of the disk \( \mathbb{D} \), under what condition does there exist a holomorphic map \( f : \mathbb{D} \to \mathbb{D} \) such that \( f(z_i) = w_i \) for all \( i \)? The necessary and sufficient condition is that the Hermitian matrix with coefficients \( \frac{1}{1-z_j z_i} \) be positive or zero. The necessity of this condition is not hard to see, but sufficiency is much more delicate (see for example [1]).

A harmonic function on the disk is a function \( h : \mathbb{D} \to \mathbb{R} \) that is the real part of a holomorphic function [3, 66, 73]. Such a function is characterized by the fact that its value at a point is the average of its values on a circle centered at the point. The dendrologic analogue is a real function defined on the vertices of a tree, whose value at each vertex is the arithmetic mean of its values on the neighboring vertices. Let us take for example a harmonic function \( h \) on the infinite tree \( A \) all of whose vertices have valence 3. Let us now suppose that \( h \) takes positive values. Evidently, if a positive number \( a \) is the average of three positive numbers, each among them is at most equal to \( 3a \). In other terms, if \( h : A \to \mathbb{R}^+ \) is harmonic and if \( x \) and \( y \) are two neighboring vertices, one has \( 1/3 \leq h(x)/h(y) \leq 3 \). We have just proved the dendrologic version of Harnack’s principle: if \( h : \mathbb{D} \to \mathbb{R}^+ \) is harmonic, the hyperbolic norm of the gradient of \( \log h \) is bounded by 1.

Now let us consider two points \( z_0 \) and \( z_1 \) in the disk. Let us define \( \text{dist}(z_0, z_1) \) as the upper bound of \( \log(h(z_0)/h(z_1)) \) over all positive harmonic functions \( h \) on the disk. Evidently, this defines a distance on the disk, homogeneous because the disk is holomorphically homogeneous. According to the characterization theorem, we find a constant multiple of the Euclidean plane or the hyperbolic plane. Of course, we have just discovered a new incarnation of the Poincaré distance! The reader may verify it for himself or consult [9].

1.10. Arrival at the boundary

We have seen that a random excursion in the disk ends almost surely on the boundary, and it is precisely on this boundary that we will end this chapter. The
points of the boundary of the Poincaré disk are not in the disk, but the founders of hyperbolic geometry quickly took note of the importance of this boundary, which they christened the absolute. There are many intrinsic definitions of the absolute (which moreover generalize to δ-hyperbolic spaces). The simplest is the following: consider the set of rays—that is, isometric embeddings \( \gamma : [0, \infty) \rightarrow \mathbb{D} \)—and identify two such rays \( \gamma_1 \) and \( \gamma_2 \) if the distance \( \text{dist}_{\text{hyp}}(\gamma_1(t), \gamma_2(t)) \) is bounded. The quotient space is by definition the absolute \( \partial \mathbb{D} \). It is not (too) difficult to endow the union \( \mathbb{D} \cup \partial \mathbb{D} \) with a topology that makes it homeomorphic to a closed disk. The action of any isometry extends canonically to the boundary [33].

Lots of things happen at the boundary: I will content myself with illustrating them by keeping a promise made above and by naturally embedding the Poincaré disk in a Hilbert space. Given two distinct points \( u \) and \( v \) of \( \partial \mathbb{D} \), there exists a unique oriented geodesic \( \gamma \subset \mathbb{D} \) that tends at infinity towards \( u \) and \( v \). Thus, the space \( \mathcal{G} \) of pairs of distinct points of the boundary can be identified with the space of oriented geodesics, and the general theory of geodesics shows that this space possesses a natural volume (or rather an area in dimension 2): this is Liouville’s theorem [6, 7]. In our case, it is not hard to identify this area. Consider the half-plane model, in which the boundary can be identified with \( \mathbb{R} \cup \{ \infty \} \). In these coordinates, the element of area is \( \omega = du 
abla / (u - v)^2 \), which one can also interpret as the cross ratio \( -[u : v : u + du : v + dv] \). If \( I \) and \( J \) are two disjoint intervals of the boundary \( \partial \mathbb{D} \), the set of pairs \( (u, v) \) such that \( u \in I \) and \( v \in J \) has an area equal to the logarithm of the cross ratio of the four endpoints of the intervals.

This being posed, one may consider the Hilbert space \( H = L^2(\mathcal{G}, \omega) \) of square-integrable functions on \( \mathcal{G} \), on which the group of isometries of the disk acts isometrically. If \( z \) is a point of the disk, we define \( \mathcal{G}_z \subset \mathcal{G} \) to be the set of pairs \( (u, v) \) such that the geodesic going from \( u \) towards \( v \) passes to the left of \( z \). We remark that the indicator function \( 1_{\mathcal{G}_z} \) of \( \mathcal{G}_z \) is not square-integrable. However, if \( z \) and \( z' \) are two points of \( \mathbb{D} \), the difference \( 1_{\mathcal{G}_z} - 1_{\mathcal{G}_{z'}} \) is square-integrable (exercise). The \( L^2 \) norm of \( 1_{\mathcal{G}_z} - 1_{\mathcal{G}_{z'}} \) is \( \text{dist}_{\text{hyp}}(z, z')^{1/2} \) (check this, without calculating!). An embedding of \( \mathbb{D} \) into \( H \) is then evident. One chooses a base point \( z_0 \) in the disk and one sends the point \( z \) to \( i(z) = \mathbb{1}_{\mathcal{G}_z} \). It is now clear that \( \|i(z) - i(z')\|^2 = \text{dist}_{\text{hyp}}(z, z') \) and that \( i \) is natural: any isometry \( f \) of \( \mathbb{D} \) naturally defines an (affine) isometry \( f \circ i \) of \( H \) such that \( i \circ f = f \circ i \).

One inconvenience of the embedding \( i \) we have just constructed is that the equality \( \|i(z) - i(z')\| = \sqrt{\text{dist}_{\text{hyp}}(z, z')} \) shows in particular that \( i \) is not differentiable, so that \( i \) is not an isometric embedding in the sense defined above! In 1932, Bieberbach constructed a natural isometric embedding of the disk into a Hilbert space [13]. Here is a modern presentation. One considers the Hilbert space of holomorphic differential forms \( \omega = f(z)\, dz \) on the disk, which are square-integrable: \( \int_\mathbb{D} \omega \wedge \overline{\omega} < \infty \). If \( z, z' \) are two fixed points of the disk, the integration of holomorphic forms along a path joining \( z \) to \( z' \) defines a linear form on \( H \) and thus, by duality, a vector \( V_{z, z'} \) in \( H \). Let \( z_0 \) be a base point of the disk. The Bieberbach embedding consists of sending \( z \) to \( j(z) = V_{z_0, z} \in H \). The naturality of \( j \) is evident because the holomorphic bijections of the disk evidently act by linear isometries of \( H \). The fact that \( j \) is differentiable is an interesting exercise. An elementary (but not very interesting) calculation shows that \( \|j(z) - j(z')\| = F(\text{dist}_{\text{hyp}}(z, z')) \) where \( F(t) = \sqrt{2\log \cosh t} \). In a neighborhood of 0, we have \( F(t) \sim t \) so that the differential of \( j \) is indeed an isometry, as claimed. On the other hand, for large
values of $f$, the difference between $F(t)$ and $\sqrt{T}$ is small and the distortion of $j$
for points distant from each other is of the same order of magnitude as for our first
embedding $i$.

This property of the natural isometric embedding into a Hilbert space has
important generalizations: a-T-menable groups, etc. (see [10] to learn more).

1.11. A few regrets...

Our excursion has not gone by so many places which would however have
deserved the detour!

I could (should?) have explained what happens on the inside of the disk rather
than limiting myself to a description of the disk, seen from the outside. The disk
is in fact a special place in which one does functional analysis [1], complex analysis
[45, 66], dynamical systems [12], number theory and modular forms [69], on which
one acts by Fuchsian groups [56], etc.

I should as well have gone farther. The disk has a numerous family. Of course,
there are versions in all dimensions (hyperbolic balls) that have analogous prop-
erties. In addition, while searching for higher dimensional Riemannian manifolds
that have strong homogeneity properties, É. Cartan founded the theory of symmetric spaces, of which he gave a magnificent classification (see for example [11, 28]).

Some of these spaces are moreover 2-homogeneous and deserve particular attention. The symmetric spaces also have combinatorial cousins: the Bruhat–Tits buildings
whose geometry contains just as many unbelievable riches [19].

And there are the spaces which are neighbors to the disk: negatively curved
manifolds, $\delta$-hyperbolic spaces and groups, etc. All this without forgetting the
infinite-dimensional hyperbolic spaces about which Gromov made the following
commentary in [39, p. 121]:

*These spaces look as cute and sexy to me as their finite-dimensional
siblings but they have been for years shamefully neglected by geometers
and algebraists alike.*

More excursions to come!

**Appendix: Sketch of a proof of the characterization theorem**

Let us consider a homogeneous metric space $(X, d)$ that is a surface, and
let us denote by $G$ its isometry group. This group has the topology of uniform
convergence on compact subsets, which in fact makes it a locally compact group
(by Ascoli’s theorem).

Hilbert’s fifth problem was solved in the middle of the 20th century by Mont-
gomery and Zippin [58]. In its final form, the result established that a locally
compact group that has a neighborhood of the identity without a non-trivial sub-
group is a Lie group, which means that this group is a differentiable manifold
and that the group structure $G \times G \to G$ is differentiable.

Replacing $X$ by its universal cover, if necessary, we may begin by assuming
that $X$ is homeomorphic to the plane or the sphere (this uses the classification of
surfaces, which was also a major event of the mathematical 20th century).
I claim that $G$ does actually contain a neighborhood of the identity without any non-trivial subgroup. For this, we use another difficult theorem, due to Kerékjártó [46, 50], according to which every compact group of homeomorphisms of the plane or the sphere is conjugate to a group of rotations (contained in $O(2)$ or $O(3)$ according to the case, the plane or the sphere). Since $O(2)$ and $O(3)$ obviously contain neighborhoods of the identity without non-trivial subgroups, the claim follows.

Following Montgomery and Zippin, the group $G$ is therefore a Lie group, which we may assume to be connected. It acts transitively on $X$ so that we can identify $X$ with $G/K$, where $K$ is the stabilizer of a point, compact of course, and therefore contained in $O(2)$ following the previous result.

The group $K$ may have dimension 0 or 1, and $G$ has dimension 2 or 3. We are thus reduced to making a list of Lie groups of dimension 2 or 3, and in the second case looking for the compact subgroups isomorphic to $O(2)$. This is not hard. Here are the possible results (always in the case where $X$ is simply connected).

a) $G$ is two-dimensional and $K$ is trivial: the space $X$ can be identified with $\mathbb{R}^2$ or with the affine group of transformations $x \mapsto ax + b$, $a > 0$.

b) $G$ is three-dimensional and $K$ is isomorphic to $O(2)$: the group $G$ can be identified with the group of orientation-preserving isometries of the sphere ($SO(3)$), the Euclidean plane, or the Poincaré disk ($PSL(2, \mathbb{R})$). In this case, the homogeneous space $X$ can be identified with the sphere, the Euclidean plane, or the Poincaré disk.

If the metric space is 2-homogeneous, case a) cannot occur and we have indeed identified, not yet the metric space, but at least its group of isometries. It is not hard to show that, under the hypothesis that $(X, d)$ is geodesic, it is in fact isometric to a constant multiple of the elliptic, Euclidean, or hyperbolic metric. To conclude, we must eliminate the hypothesis that $X$ is simply connected. This is not hard once the universal cover has been identified. This ends the sketch of the proof of the characterization theorem which, as we have seen, costs a great deal in the sense that it uses many difficult things.

If we only keep the hypothesis of homogeneity (no longer assuming 2-homogeneity), we must also consider case a). If $G \simeq \mathbb{R}^2$, it suffices to take a translation-invariant distance in the plane, for example any norm. In the same way, we may consider distances on the affine group that are invariant by translations on the left. These examples are well understood and clearly do not present the richness of the Euclidean plane, the sphere, and the Poincaré disk.
BIBLIOGRAPHY


[71] The institute for figuring: www.theiff.org/


Thanks to Aurélien Alvarez, Pierre de la Harpe, Yann Ollivier, Martine Rufin and Bruno Sévennec for their detailed readings of this text as well as for their good advice.