

5 The hyperbolic plane

5.1 Isometries

We just saw that a metric of constant negative curvature is modelled on the upper half space H with metric

$$\frac{dx^2 + dy^2}{y^2}$$

which is called the *hyperbolic plane*. This is an abstract surface in the sense that we are not considering a first fundamental form coming from an embedding in \mathbf{R}^3 , and yet it is concrete enough to be able to write down and see everything explicitly. First we consider the isometries from H to itself.

If $a, b, c, d \in \mathbf{R}$ and $ad - bc > 0$ then the Möbius transformation

$$z \mapsto w = \frac{az + b}{cz + d} \tag{10}$$

restricts to a smooth bijection from H to H with smooth inverse

$$w \mapsto z = \frac{dw - b}{-cw + a}.$$

If we substitute

$$w = \frac{az + b}{cz + d} \text{ and } dw = \left(\frac{a}{cz + d} - \frac{c(az + b)}{(cz + d)^2} \right) dz = \frac{(ad - bc)}{(cz + d)^2} dz$$

into

$$\frac{du^2 + dv^2}{v^2} = \frac{4|dw|^2}{|w - \bar{w}|^2}$$

we get

$$\frac{4(ad - bc)^2 |dz|^2}{|(az + b)(c\bar{z} + d) - (a\bar{z} + b)(cz + d)|^2} = \frac{4(ad - bc)^2 |dz|^2}{|(ad - bc)(z - \bar{z})|^2} = \frac{4|dz|^2}{|z - \bar{z}|^2} = \frac{dx^2 + dy^2}{y^2}.$$

Thus this Möbius transformation is an isometry from H to H . So is the transformation $z \mapsto -\bar{z}$, and hence the composition

$$z \mapsto \frac{b - a\bar{z}}{d - c\bar{z}} \tag{11}$$

is also an isometry from H to H . In fact (10) and (11) give all the isometries of H , as we shall see later.

In **4.3** we saw that the unit disc D with the metric

$$\frac{du^2 + dv^2}{(1 - u^2 - v^2)^2}$$

is isometric to H , so any statements about H transfer also to D . Sometimes the picture is easier in one model or the other. The isometries $f : D \rightarrow D$ of the unit disc model of the hyperbolic plane are also Möbius transformations, if they preserve orientations, or compositions of Möbius transformations with $z \mapsto \bar{z}$ if they reverse orientations. The Möbius transformations which map D to itself are those of the form

$$z \mapsto w = e^{i\theta} \left(\frac{z - a}{1 - \bar{a}z} \right)$$

where $a \in D$ and $\theta \in \mathbf{R}$. They are isometries because substituting for w and

$$dw = e^{i\theta} \frac{(1 - |a|^2)}{(1 - \bar{a}z)^2} dz$$

in $4(1 - |w|^2)^{-2}|dw|^2$ gives $4(1 - |z|^2)^{-2}|dz|^2$.

Notice that the group $\text{Isom}(H)$ of isometries of H acts transitively on H because if $a + ib \in H$ then $b > 0$ so the transformation

$$z \mapsto bz + a$$

is an isometry of H which takes i to $a + ib$. Similarly the group $\text{Isom}(D)$ of isometries of D acts transitively on D since if $a \in D$ then the isometry

$$z \mapsto \frac{z - a}{1 - \bar{a}z}$$

maps a to 0. Notice also that the subgroup of $\text{Isom}(D)$ consisting of those isometries which fix 0 contains all the rotations

$$z \mapsto e^{i\theta} z$$

about 0 as well as $z \mapsto \bar{z}$.

5.2 Geodesics

The hyperbolic plane is a case where the geodesic equations can be easily solved: since $E = G = 1/y^2$ and $F = 0$, and these are independent of x , the first geodesic equation

$$\frac{d}{ds}(Eu' + Fv') = \frac{1}{2}(E_u u'^2 + 2F_u u'v' + G_u v'^2)$$

becomes

$$\frac{d}{ds} \left(\frac{x'}{y^2} \right) = 0$$

and so

$$x' = cy^2. \tag{12}$$

We also know that parametrization is by arc length in these equations so

$$\frac{x'^2 + y'^2}{y^2} = 1 \tag{13}$$

If $c = 0$ we get $x = \text{const.}$, which is a vertical line. Suppose $c \neq 0$, then from (12) and (13) we have

$$\frac{dy}{dx} = \sqrt{\frac{y^2 - c^2 y^4}{c^2 y^4}}$$

or

$$\frac{cydy}{\sqrt{1 - c^2 y^2}} = dx$$

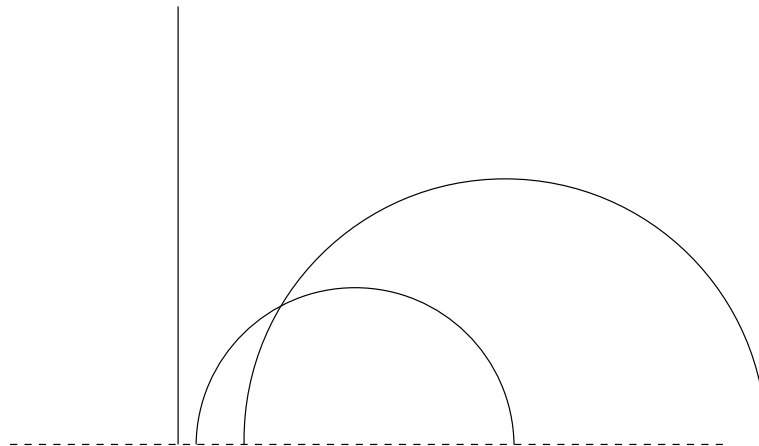
which integrates directly to

$$-c^{-1} \sqrt{1 - c^2 y^2} = x - a$$

or

$$(x - a)^2 + y^2 = 1/c^2$$

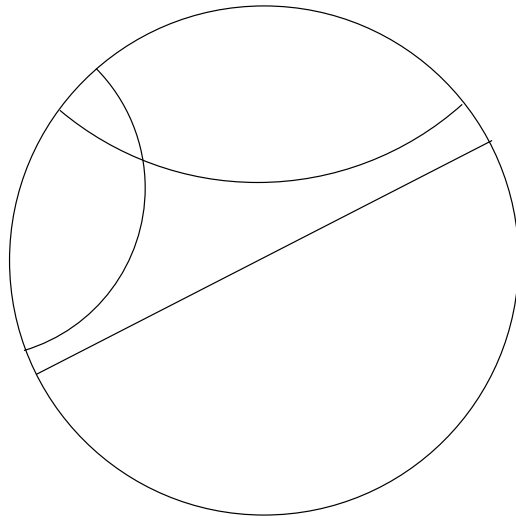
which is a semicircle centred on the real axis.



The isometry from H to D given by

$$w \mapsto z = \frac{w - i}{w + i}$$

takes geodesics to geodesics (since it is an isometry) and it is the restriction to H of a Möbius transformation $\mathbf{C} \cup \{\infty\} \rightarrow \mathbf{C} \cup \{\infty\}$ which takes circles and lines to circles and lines, preserves angles and maps the real axis to the unit circle in \mathbf{C} . It therefore follows that the geodesics in D are the circles and lines in D which meet the unit circle at right angles.



Using geodesics we can now show that any isometry is a Möbius transformation as above. So suppose that $F : D \rightarrow D$ is an isometry. Take a Möbius isometry G taking $F(0)$ to 0, then we need to prove that GF is Möbius. This is an isometry fixing 0, so it takes geodesics through 0 to geodesics through 0. It preserves angles, so acts on those geodesics by a rotation or reflection. It also preserves distance so it takes a point on a geodesic a distance r from the origin to another point at the same distance. However, as we noted above, each rotation $R : z \mapsto e^{i\theta}z$ is a Möbius isometry, so composing with this we see that $RGF = 1$ and $F = (RG)^{-1}$ is a Möbius isometry.

5.3 Angles and distances

Hyperbolic angles in H and in D are the same as Euclidean angles, since their first fundamental forms satisfy $E = G$ and $F = 0$. Distances between points are given by the lengths of geodesics joining the points. Since the interval $(-1, 1)$ is a geodesic in the unit disc D , the distance from 0 to any $x \in (0, 1)$ is given by the hyperbolic length of the line segment $[0, x]$, which is

$$\int_0^x \sqrt{Eu'^2 + 2Fu'v' + Gv'^2} dt = \int_0^x \frac{dt}{1-t^2} = 2 \tanh^{-1} x$$

where $u(t) = t$ and $v(t) = 0$ and $E = G = (1 - u^2 - v^2)^{-2}$ and $F = 0$. Given any $a, b \in D$ we can choose $\theta \in \mathbf{R}$ such that

$$e^{i\theta} \frac{b - a}{1 - \bar{a}b} = \left| \frac{b - a}{1 - \bar{a}b} \right|$$

is real and positive, so its distance from 0 is

$$2 \tanh^{-1} \left| \frac{b - a}{1 - \bar{a}b} \right|.$$

Since the isometry

$$z \mapsto e^{i\theta} \frac{z - a}{1 - \bar{a}z}$$

preserves distances and takes a to 0 and b to $e^{i\theta}(b - a)/(1 - \bar{a}b)$, it follows that the hyperbolic distance from a to b in D is

$$d_D(a, b) = 2 \tanh^{-1} \left| \frac{b - a}{1 - \bar{a}b} \right|.$$

We can work out hyperbolic distances in H in a similar way by first calculating the distance from i to λi for $\lambda \in [1, \infty)$ as the length of the geodesic from i to λi given by the imaginary axis, which is

$$\int_1^\lambda \frac{dt}{t} = \log \lambda,$$

and then given $a, b \in H$ finding an isometry of H which takes a to i and b to λi for some $\lambda \in [1, \infty)$. Alternatively, since we have an isometry from H to D given by

$$w \mapsto z = \frac{w - i}{w + i},$$

the hyperbolic distance between points $a, b \in H$ is equal to the hyperbolic distance between the corresponding points $(a - i)/(a + i)$ and $(b - i)/(b + i)$ in D , which is

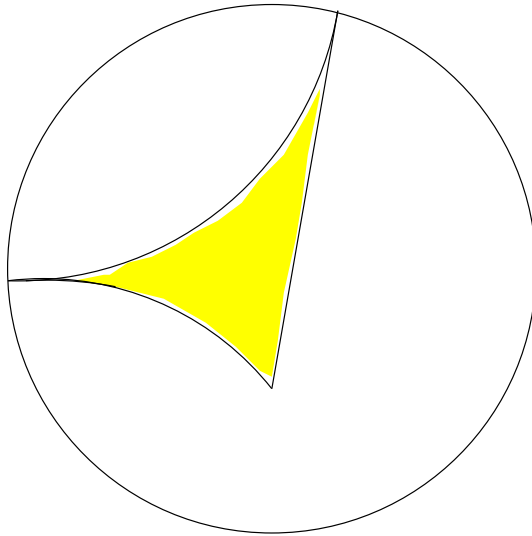
$$d_H(a, b) = d_D\left(\frac{a - i}{a + i}, \frac{b - i}{b + i}\right) = 2 \tanh^{-1} \left| \frac{(b - i)(a + i) - (a - i)(b + i)}{(a + i)(b + i) - (a - i)(b - i)} \right| = 2 \tanh^{-1} \left| \frac{b - a}{b - \bar{a}} \right|.$$

5.4 Hyperbolic triangles

A hyperbolic triangle Δ is given by three distinct points in H or D joined by geodesics. We see immediately from Gauss-Bonnet that the sum of the angles of a triangle is given by

$$A + B + C = \pi - \text{Area}(\Delta).$$

We can also consider hyperbolic triangles which have one or more vertices ‘at infinity’, i.e. on the boundary of H or D . These triangles are called *asymptotic*, *doubly (or bi-) asymptotic* and *triply (or tri-) asymptotic*, according to the number of vertices at infinity. The angle at a vertex at infinity is always 0, since all geodesics in H or D meet the boundary at right angles.



Theorem 5.1 (The cosine rule for hyperbolic triangles) *If Δ is a hyperbolic triangle in D with vertices at a, b, c and*

$$\alpha = d_D(b, c), \quad \beta = d_D(a, c) \text{ and } \gamma = d_D(a, b)$$

then

$$\cosh \gamma = \cosh \alpha \cosh \beta - \sinh \alpha \sinh \beta \cos \theta$$

where θ is the internal angle of Δ at c .

Proof: Because the group of isometries of D acts transitively on D we can assume that $c = 0$. Moreover, since the rotations $z \mapsto e^{i\phi}z$ are isometries which fix 0, we can also assume that a is real and positive. Then $\beta = 2 \tanh^{-1}(a)$ so

$$a = \tanh(\beta/2)$$

and similarly

$$b = e^{i\theta} \tanh(\alpha/2)$$

while

$$\tanh(\gamma/2) = \left| \frac{b - a}{1 - \bar{a}b} \right|.$$

Recall that

$$\frac{1 + \tanh^2(\gamma/2)}{1 - \tanh^2(\gamma/2)} = \cosh(\gamma)$$

so

$$\cosh(\gamma) = \frac{|1 - \bar{a}b|^2 + |b - a|^2}{|1 - \bar{a}b|^2 - |b - a|^2} = \frac{(1 + |a|^2)(1 + |b|^2) - 2(\bar{a}b + a\bar{b})}{(1 - |a|^2)(1 - |b|^2)}.$$

Now

$$\frac{1 + |a|^2}{1 - |a|^2} = \frac{1 + \tanh^2(\beta/2)}{1 - \tanh^2(\beta/2)} = \cosh\beta$$

as above, and similarly

$$\frac{1 + |b|^2}{1 - |b|^2} = \cosh\alpha$$

while

$$\frac{2(\bar{a}b + a\bar{b})}{(1 - |a|^2)(1 - |b|^2)} = \frac{2\tanh(\alpha/2)\tanh(\beta/2)(e^{i\theta} + e^{-i\theta})}{\operatorname{sech}^2(\alpha/2)\operatorname{sech}^2(\beta/2)} = \sinh\alpha \sinh\beta \cos\theta.$$

This completes the proof. \square

Theorem 5.2 (The sine rule for hyperbolic triangles) Let Δ be a hyperbolic triangle in D with internal angles A, B, C at vertices a, b, c and

$$\alpha = d_D(b, c), \quad \beta = d_D(a, c) \text{ and } \gamma = d_D(a, b).$$

Then

$$\frac{\sin A}{\sinh\alpha} = \frac{\sin B}{\sinh\beta} = \frac{\sin C}{\sinh\gamma}.$$

Proof: Two alternatives approaches:

1) Use the cosine rule to find an expression for $\sinh^2\alpha \sinh^2\beta \sin^2C$ in terms of $\cosh\alpha$, $\cosh\beta$ and $\cosh\gamma$ which is symmetric in α , β and γ , and deduce that

$$\sinh^2\alpha \sinh^2\beta \sin^2C = \sinh^2\alpha \sinh^2\gamma \sin^2B = \sinh^2\gamma \sinh^2\beta \sin^2A.$$

2) First prove that if $C = \pi/2$ then $\sin A \sinh\gamma = \sinh\alpha$ by applying the cosine rule to Δ in two different ways. Then deduce the result in general by dropping a perpendicular from one vertex of Δ to the opposite side. \square

Gauss-Bonnet and its limits give the following:

Theorem 5.3 (Areas of hyperbolic triangles)

- (i) *The area of a triply asymptotic hyperbolic triangle Δ is π .*
- (ii) *The area of a doubly asymptotic hyperbolic triangle Δ with internal angle θ is $\pi - \theta$.*
- (iii) *The area of an asymptotic hyperbolic triangle Δ with internal angles θ and ϕ is $\pi - \theta - \phi$.*
- (iv) *The area of a hyperbolic triangle Δ with internal angles θ , ϕ and ψ is $\pi - \theta - \phi - \psi$.*

5.5 Non-Euclidean geometry

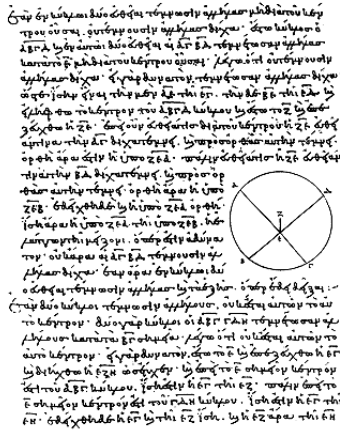
As we see above, the analogy between Euclidean geometry and its theorems and the geometry of the hyperbolic plane is very close, so long as we replace lines by geodesics, and Euclidean isometries (translations, rotations and reflections) by the isometries of H or D . In fact it played an important historical role.

For centuries, Euclid's deduction of geometrical theorems from self-evident common notions and postulates was thought not only to represent a model of the physical space in which we live, but also some absolute logical structure. One postulate caused some problems though – was it really self-evident? Did it follow from the other axioms? This is how Euclid phrased it:

“That if a straight line falling on two straight lines makes the interior angle on the same side less than two right angles, the two straight lines if produced indefinitely, meet on that side on which the angles are less than two right angles”.

Some early commentators of Euclid's *Elements*, like Posidonius (1st Century BC), Geminus (1st Century BC), Ptolemy (2nd Century AD), Proclus (410 - 485) all felt that the parallel postulate was not sufficiently evident to accept without proof.

Here is a page from a medieval edition of Euclid dating from the year 888. It is handwritten in Greek. The manuscript, contained in the Bodleian Library, is one of the earliest surviving editions of Euclid.



The controversy went on and on with Greek and Islamic mathematicians puzzling over it. Johann Lambert (1728-1777) realized that if the parallel postulate did not hold then the angles of a triangle add up to less than 180° , and that the deficit was the area. He found this worrying in many ways, not least because it says that there is an absolute scale – no distinction between similar and congruent triangles. Finally Janos Bolyai (1802-1860) and Nikolai Lobachevsky (1793-1856) discovered non-Euclidean geometry simultaneously. It satisfies all of Euclid’s axioms except the parallel postulate, and we shall see that it is the geometry of H or D that we have been studying.

Bolyai became interested in the theory of parallel lines under the influence of his father Farkas, who devoted considerable energy towards finding a proof of the parallel postulate without success. He even wrote to his son:

“I entreat you, leave the doctrine of parallel lines alone; you should fear it like a sensual passion; it will deprive you of health, leisure and peace – it will destroy all joy in your life.”

Another relevant figure in the discovery was Carl Friedrich Gauss (1777-1855), who as we have seen developed the differential geometry of surfaces. He was the first to consider the possibility of a geometry denying the parallel postulate. However, for fear of being ridiculed he kept his work unpublished, or maybe he never made the connection with the curvature of real world surfaces and the Platonic ideal of axiomatic geometry. Anyway, when he read Janos Bolyai’s work he wrote to Janos’s father:

“If I commenced by saying that I must not praise this work you would certainly be surprised for a moment. But I cannot say otherwise. To praise it, would be to praise

myself. Indeed the whole contents of the work, the path taken by your son, the results to which he is led, coincide almost entirely with my meditations, which have occupied my mind partly for the last thirty or thirty-five years."

Euclid's axioms were made rigorous by Hilbert. They begin with undefined concepts of

- "point"
- "line"
- "lie on" (a point **lies on** a line)
- "betweenness"
- "congruence of pairs of points"
- "congruence of pairs of angles".

Euclidean geometry is then determined by logical deduction from the following axioms:

EUCLID'S AXIOMS

I. AXIOMS OF INCIDENCE

1. Two points have one and only one straight line in common.
2. Every straight line contains a least two points.
3. There are at least three points not lying on the same straight line.

II. AXIOMS OF ORDER

1. Of any three points on a straight line, one and only one lies between the other two.
2. If A and B are two points there is at least one point C such that B lies between A and C .
3. Any straight line intersecting a side of a triangle either passes through the opposite vertex or intersects a second side.

III. AXIOMS OF CONGRUENCE

1. On a straight line a given segment can be laid off on either side of a given point (the segment thus constructed is congruent to the give segment).

2. If two segments are congruent to a third segment, then they are congruent to each other.
3. If AB and $A'B'$ are two congruent segments and if the points C and C' lying on AB and $A'B'$ respectively are such that one of the segments into which AB is divided by C is congruent to one of the segments into which $A'B'$ is divided by C' , then the other segment of AB is also congruent to the other segment of $A'B'$.
4. A given angle can be laid off in one and only one way on either side of a given half-line; (the angle thus drawn is congruent to the given angle).
5. If two sides of a triangle are equal respectively to two sides of another triangle, and if the included angles are equal, the triangles are congruent.

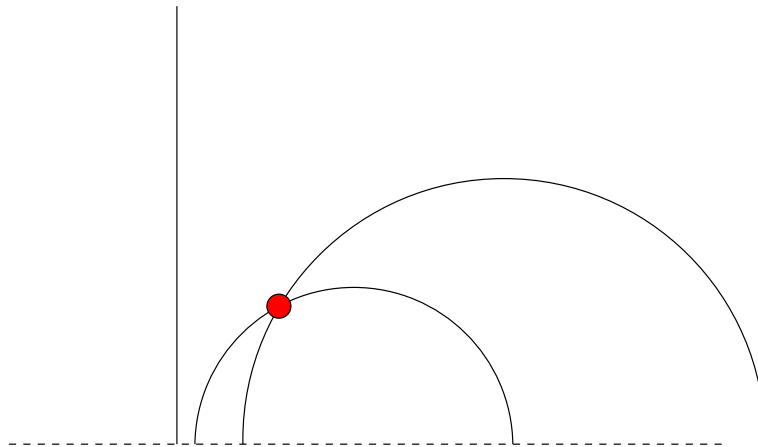
IV. AXIOM OF PARALLELS

Through any point not lying on a straight line there passes one and only one straight line that does not intersect the given line.

V. AXIOM OF CONTINUITY

1. If AB and CD are any two segments, then there exists on the line AB a number of points A_1, \dots, A_n such that the segments $AA_1, A_1A_2, \dots, A_{n-1}A_n$ are congruent to CD and such that B lies between A and A_n

Clearly H does not satisfy the Axiom of Parallels:



The fact that hyperbolic geometry satisfies all the axioms except the parallel postulate is now only of historic significance and the reader is invited to do all the checking. Often one model is easier than another. Congruence should be defined through the action of the group of isometries.

5.6 Complex analysis and the hyperbolic plane

The intricate metric structure of the hyperbolic plane – geodesics, triangles and all – is actually determined purely by the holomorphic functions on it, so we could also think of hyperbolic geometry as a branch of complex analysis. Here is the theorem that makes it work:

Theorem 5.4 *Any holomorphic homeomorphism $f : D \rightarrow D$ is an isometry of the hyperbolic metric.*

Proof: The argument follows Schwartz's lemma. By applying an isometry we can assume that $f(0) = 0$, and since the image of f is D , we have $|f(z)| < 1$ if $z \in D$. Now since $f(0) = 0$, $f_1(z) = f(z)/z$ is holomorphic and applying the maximum principle to a disc of radius $r < 1$ we get

$$|f_1(z)| \leq \frac{1}{r}$$

and in the limit as $r \rightarrow 1$, $|f_1(z)| \leq 1$ or equivalently

$$|f(z)| \leq |z|.$$

Since f is a homeomorphism, its inverse satisfies the same inequality so

$$|z| \leq |f(z)|$$

and $|f_1(z)| = 1$ everywhere. Since this is true at an interior point the function must be a constant c so $f(z) = cz$ and since $|f(z)| = |z|$,

$$f(z) = e^{i\theta} z$$

which is an isometry. □

In fact a similar result holds for \mathbf{C} :

Theorem 5.5 *Any holomorphic homeomorphism $f : \mathbf{C} \rightarrow \mathbf{C}$ is of the form $f(z) = az + b$ with $a \neq 0$.*

If $|a| = 1$ this is an isometry of the Euclidean metric $dx^2 + dy^2$. The extra scaling $z \mapsto \lambda z$ is what gives rise in classical geometrical terms to similar but non-congruent triangles.

Proof: For $|z| > R$ consider the function $g(z) = f(1/z)$. Suppose g has an essential singularity at $z = 0$. Then the Casorati-Weierstrass theorem (Exercise 17.5 in [4]) tells us that $g(z)$ gets arbitrarily close to any complex number if z is small enough, and in particular to values in the image of $\{z : |z| \leq R\}$ under f . But we assumed f was bijective, which is a contradiction. It follows that g has at most a pole at infinity and so $f(z)$ must be a polynomial of some degree k .

However the equation $f(z) = c$ then has k solutions for most values of c , and again since f is bijective we must have $k = 1$ and

$$f(z) = az + b.$$

□

For completeness, we add the following

Theorem 5.6 *Any holomorphic homeomorphism f of the Riemann sphere to itself is a Möbius transformation $z \mapsto (az + b)/(cz + d)$.*

Proof: By using a Möbius transformation we can assume that $f(\infty) = \infty$ and then the previous theorem tells us that $f(z) = az + b$. □

These results are all about the complex plane and its subsets. In fact hyperbolic geometry has an important role to play in the study of compact Riemann surfaces. Recall that local holomorphic coordinates on a Riemann surface are related by holomorphic transformations and these preserve angles. Given two smooth curves on a Riemann surface, it makes good sense to define their angle of intersection and this is called a *conformal structure*. A metric also defines angles so we can consider metrics compatible with the conformal structure of a Riemann surface. In a local coordinate z such a metric is of the form

$$f dz d\bar{z} = f(dx^2 + dy^2).$$

The remarkable result is the following *uniformization* theorem:

Theorem 5.7 *Every closed Riemann surface X has a metric of constant Gaussian curvature compatible with its conformal structure.*

Note that by the Gauss-Bonnet theorem $K > 0$ implies $\chi(X) > 0$, i.e. X is a sphere, $K = 0$ implies $\chi(X) = 0$, i.e. X is a torus, and $K < 0$ gives $\chi(X) < 0$.

Proof: The proof is a corollary of a difficult theorem called the Riemann mapping theorem. Recall that a space is *simply-connected* if it is connected and every closed path can be shrunk to a point. The Riemann mapping theorem (proved by Poincaré and Koebe) says that every simply-connected Riemann surface is holomorphically homeomorphic to either the Riemann sphere, \mathbf{C} or H .

If X is any reasonable topological space, one can form its *universal covering space* \tilde{X} (see [3]) which is simply connected and has

- a projection $p : \tilde{X} \rightarrow X$
- every point $x \in X$ has a neighbourhood V such that $p^{-1}(V)$ consists of a disjoint union of open sets each of which is homeomorphic to V by p
- there is a group π of homeomorphisms of \tilde{X} such that $p(gy) = p(y)$, so that π permutes the different sheets in $p^{-1}(V)$.
- no element of π apart from the identity has a fixed point
- X can be identified with the space of orbits of π acting on \tilde{X} .

The standard example of this is $X = S^1$, $\tilde{X} = \mathbf{R}$, $p(t) = e^{it}$ and $\pi = \mathbf{Z}$ acting by $t \mapsto t + 2n\pi$. It is easy to see that the universal covering of a Riemann surface is a Riemann surface, so applying the Riemann mapping theorem we see that \tilde{X} is either the Riemann sphere, \mathbf{C} or H .

So consider the cases:

- If \tilde{X} is the sphere S , it is compact and so $p : \tilde{X} \rightarrow X$ has only a finite number k of sheets. By counting vertices, edges and faces it is clear that $\chi(\tilde{X}) = k\chi(X)$. Since $\chi(S) = 2$, we must have $k = 1$ or 2 , but if the latter $\chi(X) = 1$ which is not of the allowable form $2 - 2g$ for an orientable surface and a Riemann surface *is* orientable. So it is only the Riemann sphere in this case.
- If $\tilde{X} = \mathbf{C}$, we appeal to Theorem 5.5. The group π of covering transformations is holomorphic and so each element is of the form $z \mapsto az + b$. But π has no fixed points, so $az + b = z$ has no solution which means that $a = 1$. the transformations $z \mapsto z + b$ are just translations and are isometries of the metric $dx^2 + dy^2$ which has $K = 0$.
- If $\tilde{X} = H$, then from Theorem 5.4, the action of π preserves the hyperbolic metric.

□

So we see that these abstract metrics have a role to play in the study of Riemann surfaces – a long long way from surfaces in \mathbf{R}^3 .