Hyperbolic geometry
MA 448

Caroline Series

With assistance from Sara Maloni

Figures by Sara Maloni and Khadija Farooq

2008; Revised 2010

Last edited: January 4, 2013

Contents

Bibliography

[*] denotes books closest to this course.

[1] L. Ahlfors *Complex analysis*, McGraw Hill (1953).

[2] D. Mumford, C. Series and D. Wright *Indra's Pearls: The vision of Felix Klein*, Cambridge University Press (2002).

[3] G. Jones and D. Singerman *Complex functions*, Cambridge University Press (1987).

[4] J. Milnor *Dynamics in one complex variable*, Princeton University Press (1931).

[5] J. Stillwell *Mathematics and its history*, Springer-Verlag (1989).

[6] *J. Anderson *Hyperbolic Geometry*, Springer-Verlag (2nd edition 2005).

[7] *S. Katok *Fuchsian groups*, University of Chicago Press (1992).

[8] *S. Stahl *The Poincaré half-plane; a gateway to modern geometry*, Jones and Bartlett Publishers (1993).

[9] M. Reid and B. Szendroi *Geometry and Topology*, Cambridge Univ. Press (2005).

[10] A. F. Beardon *The Geometry of Discrete Groups*, Springer-Verlag (1983).

[11] M. do Carmo. *Riemannian Geometry*, Birkhäuser (1992).

[12] L. R. Ford *Automorphic functions*, Chelsea Publishing Company (1929).

[13] J. Lehner *Discontinuous Groups and Automorphic Functions*, American Mathematical Society (1964).

[14] W. P. Thurston, edited by Silvio Levy *Three-dimensional geometry and topology*, Princeton Univ. Press (1997).

[15] W. P. Thurston *The geometry and topology of three-manifolds*, Princeton University Mathematics Department (1979), lecture notes.

[16] F.Bonahon *Low dimensional Geometry*, New book shortly to appear; available at `http://almaak.usc.edu/ fbonahon/Research/Preprints/Preprints.html`

[17] D. B. A. Epstein *Geometric structures on Manifolds*, The Mathematical Intelligencer 14, Number 1, (1982), 5-16.

[18] D. Hilbert and S. Cohn-Vossen *Geometry and the Imagination*, Chelsea Publishing Company, (1952).

[19] M. A. Armstrong *Basic Topology*, Springer-Verlag, (1983).

[20] A. Hatcher *Algebraic Topology*, Cambridge Univ. Press, (2002).

Introduction

0.1 Introduction

What is hyperbolic geometry and why study it?

Spaces of constant curvature Hyperbolic (also called non-Euclidean) geometry is the study of geometry on spaces of constant negative curvature.

In dimension 2, surfaces of constant curvature are distinguished by whether their curvature K is positive, zero or negative. Given such a surface in $\mathbb{R}^3$, if $K > 0$ all points of such a surface X are on one side of the tangent plane any the point; if $K = 0$ then there is always a line in X contained in the tangent plane; and if $K < 0$ there are points of X on both sides of any tangent plane. If the space is simply connected and if $K > 0$, then it is a sphere (with $K = 1/\text{radius}$); if $K = 0$ it is the Euclidean plane; and if $K < 0$ it is the hyperbolic plane, also called 2-dimensional hyperbolic space. Such surfaces look the same at every point and in every direction and so ought to have lots of symmetries.

The geometry of the sphere and the plane are familiar; hyperbolic geometry is the geometry of the third case.

Hyperbolic space has many interesting features; some are similar to those of Euclidean geometry but some are quite different. In particular it has a very rich group of isometries, allowing a huge variety of crystallographic symmetry patterns. This makes the geometry both rigid and flexible at the same time. Its properties and symmetries are closely related to tree-like growth patterns and fractals, suggestive of many natural biological objects like ferns and trees.

Historical importance Another way to describe hyperbolic geometry is to say that is the geometry of space in which Euclid's parallel axiom fails. One way to state the parallel axiom is that for every line L, and point P not on L, there is a unique line L' through P which does not meet L, even if extended infinitely far in both directions.

Historically, hyperbolic geometry had enormous importance. Attempts to deduce the parallel axiom from Euclid's other axioms led to many developments within Euclidean geometry. Realisation slowly dawned that the reason that all attempts to deduce the parallel axiom failed, was that one could develop a consistent system of geometry which was not Euclidean, in which the parallel axiom failed while the other axioms remained true. This had enormous scientific and philosophical importance because it showed that mathematically, there was nothing absolute about Euclidean space.

Hyperbolic geometry and surfaces Any topological surface can be endowed with a *geometric structure*. This means that one can find a metric on the surface which in small regions looks like one of the three above types of geometry, and such that the 'overlap maps' (c.f. the definition of a manifold to understand this) are isometries of the appropriate geometry. Thus a cylinder or a torus carries a Euclidean structure while, as we shall see later in the course, every surface of negative Euler characteristic (in particular every closed surface of genus at least 2) carries a hyperbolic structure. In fact, in all cases except the sphere, the surfaces in question carry not one but many such metrics.

Being able to put a hyperbolic metric on a surface allows one to study many features of the surface very precisely. This applies especially to the study of diffeomorphisms on surfaces, where hyperbolic geometry plays a crucial role.

Revolutionary work by William Thurston in the 1980's opened up the possiblity of a similar description of 3-manifolds. He showed that many 3-manifolds are endowed with a natural geometrical structures of one of a few kinds, of which by far the most common is hyperbolic. He conjectured that any topological 3-manifold could be cut into pieces on the basis of topological information alone, such that each piece carries one of 8 special geometries, among them hyperbolic. This *geometrisation conjecture* has now almost certainly been proved in consequence of the work of Perelman.

Thurston's insights led to a great revolution in the study of hyperbolic 3-manifolds, making it a large and very active area of mathematics in which spectacular progress has been made in recent years. Besides Thurston, some of the leading names are: Agol, Bonahon, Brock, Canary, Gabai, Masur, Minsky, and Otal.

Connection with other parts of mathematics Hyperbolic geometry is closely connected to many other parts of mathematics. Here are some: differential geometry, complex analysis, topology, dynamical systems including complex dynamics and ergodic theory, relativity, Diophantine approximation and number theory, geometric group theory, Riemann surfaces, Teichmüller theory.

In this course we will study two closely related models of hyperbolic geometry, that is, spaces with constant curvature -1. These are the upper half plane $\mathbb{H} = \{z \in \mathbb{C} : \Im z > 1\}$ with the metric $ds = |dz|/\Im z$ and the unit disk $\mathbb{D} = \{z \in \mathbb{C} : |z| < 1\}$ with the metric $ds = \dfrac{2|dz|}{(1-|z|^2)}$. (Here $|dz| := \sqrt{dx^2 + dy^2}$.)

Exercise 0.1. *From differential geometry, the curvature K of a Riemann surface with metric $ds^2 = g_1 dx_1^2 + g_2 dx_2^2$ is given by the formula*

$$K = -\frac{1}{\sqrt{g_1 g_2}}\left(\frac{\partial}{\partial x_1}\Big(\frac{1}{\sqrt{g}_1}\frac{\partial\sqrt{g}_2}{\partial x_1}\Big) + \Big(\frac{\partial}{\partial x_2}\frac{1}{\sqrt{g}_2}\frac{\partial\sqrt{g}_1}{\partial x_2}\Big)\right).$$

Verify that the curvature of the above metrics in both $\mathbb{H}$ and $\mathbb{D}$ is -1.

0.2 Models of hyperbolic geometry

Unlike the situation in spherical geometry, it is impossible to embed an infinite simply connected surface of constant negative curvature isometrically into Euclidean 3-space. (This is a theorem of Hilbert, proved in 1901).[1] Think of a kale leaf which crinkles up more and more as you go towards its edge. Thus to visualize hyperbolic geometry, we have to resort to a model.

[1]Perhaps this is the reason for the historically late development of hyperbolic geometry.

Figure 1: A crinkled kale leaf

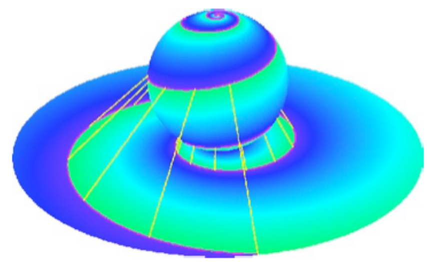

Figure 2: Stereographic projection of the sphere to the plane

You can think of this as similar to the projection of geometry on the surface of a sphere by stereographic projection (or other map projection) onto the plane. The projection introduces distortion so that either distance or angles measurements, or both, do not look correct. Like stereographic projection, the two models we will study preserve angles (are conformal) but massively distort distance.

We are going to study two such projections or models which are in fact almost equivalent: the upper half plane model $\mathbb{H} = \{z \in \mathbb{C} : \Im z > 1\}$ with the metric $ds = |dz|/\Im z$ (also known as the Lobachevsky plane) and the unit disk $\mathbb{D} = \{z \in \mathbb{C} : |z| < 1\}$ with the metric $ds = \dfrac{2|dz|}{(1 - |z|^2)}$ (also known as the Poincaré disk). In both models:

Lines are arcs of circles (or lines) which meet $\partial\mathbb{D}$ or $\partial\mathbb{H}$ orthogonally. Notice that in $\mathbb{H}$ such arcs must be semicircles with centres on $\mathbb{R}$; in $\mathbb{D}$ such an arc never has its centre on $\partial\mathbb{D}$.

Angles The usual Euclidean angle.

Distance The *infinitesimal form* of the metric on $\mathbb{H}$ is $ds = |dz|/\Im z$ and on $\mathbb{D}$ is $ds = 2|dz|/(1 - |z|^2)$.

The actual formula for distance between two points can be expressed in terms of *cross-ratio*. If z_1, z_2, z_3, z_4 are distinct points in $\mathbb{C}$ then their cross-ratio is defined as

$$[z_1, z_2, z_3, z_4] = \frac{z_1 - z_2}{z_1 - z_3} \cdot \frac{z_4 - z_3}{z_4 - z_2}.$$

Now let P, Q be two points in $\mathbb{D}$ or $\mathbb{H}$. You can prove by Euclidean geometry that there is a *unique* (hyperbolic) line joining P to Q. Let P', Q' be the points where the extension of this line meets $\partial\mathbb{D}$ or $\partial\mathbb{H}$, so that the order along the line is P', P, Q, Q'. Then:

$$d_{\mathbb{D}}(P,Q) = \log[P',Q,P,Q'] \quad \text{and} \quad d_{\mathbb{H}}(P,Q) = \log[P',Q,P,Q']$$

Isometries Poincaré (1881) realised that the *isometries* or distance preserving bijections of these two models are exactly the *linear fractional transformations* or *Möbius maps* $z \mapsto \dfrac{az+b}{cz+d}$ (where $a,b,c,d \in \mathbb{C}$) which preserve $\mathbb{D}$ or $\mathbb{H}$ respectively. This makes it particularly easy to study and compute with such maps. It turns out that they are also the set of all *conformal automorphisms* of $\mathbb{D}$ or $\mathbb{H}$. We shall review linear fractional transformations in the next chapter.

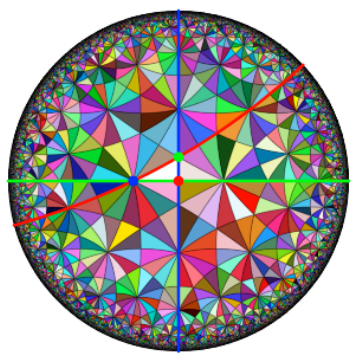

Figure 3: A tiling of the hyperbolic plane, made by hyperbolically reflecting in the sides of the white triangle whose sides are marked with red, blue and green dots.

0.3 Other models of hyperbolic geometry

Here we just give very brief descriptions. Good references for this are [9, 14]

The hemisphere model Think of the unit disk $\mathbb{D}$ in $\mathbb{R}^3$ as contained in the 'horizontal' plane $x_0 = 0$. Let S^2 denote the unit sphere in $\mathbb{R}^3$. Stereographically project $\mathbb{D}$ from the south pole $(-1,0,0)$ onto the northern hemisphere of S^2. This transfers the geometry of $\mathbb{D}$ to geometry on the hemisphere. The map is conformal or angle preserving (because stereoegraphic projection can be described as a product of inversions). Geodesics are 'vertical' semicircles with centres on the plane $x_0 = 0$ and orthogonal to $\partial\mathbb{D}$.

The hyperboloid model In this model, hyperbolic space is geometry on a 'sphere of radius i'. Consider $\mathbb{R}^3$ with the indefinite Lorentz metric

$$ds^2 = -dx_0^2 + dx_1^2 + dx_2^2,$$

associated to the quadratic form $Q((x_0, x_1, x_2)) = -x_0^2 + x_1^2 + x_2^2$. The length of a vector $\mathbf{x} = (x_0, x_1, x_2)$ is $\sqrt{Q(\mathbf{x})}$.

The sphere of radius i about the origin is the set

$$H = \{(x_0, x_1, x_2) : Q(\mathbf{x}) = -1\}.$$

Restricted to H, the metric ds^2 becomes a positive definite Riemanian metric: you can check that any tangent vector to H has positive length. (Suppose $\mathbf{x} \in H$ so that $Q(\mathbf{x}, \mathbf{x}) = -\mathbf{1}$. Differentiating along any path $\mathbf{x(t)} \in \mathbf{H}$ gives $Q(\mathbf{x}, \dot{\mathbf{x}}) = \mathbf{0}$. In other words, a vector $\dot{\mathbf{x}}$ in the tangent space at $\mathbf{x}$ is in the orthogonal complement of $\mathbf{x}$ with respect to Q, and this orthogonal complement has signature $+2$.) H is a hyperboloid with two components depending on the sign of x_0. Hyperbolic space is modelled on one sheet of this hyperboloid, usually the upper sheet H^+ where $x_0 > 0$. (Thinking of the direction $(1, 0, 0)$ as 'vertical'.) The hyperbolic distance between two points $\mathbf{x}, \mathbf{y}$ is given by $\cosh d(\mathbf{x}, \mathbf{y}) = -Q(\mathbf{x}, \mathbf{y})$. Geodesics in H^+ are exactly the intersection of planes through the origin with H^+. There is a nice explanation of all this in [9], starting with the easier case of 'one dimensional hyperbolic space' and the Lorenz metric on $\mathbb{R}^2$.

The set of linear maps which preserve Q, (ie maps $A : \mathbb{R}^3 \longrightarrow \mathbb{R}^3$ such that $Q(A\mathbf{x}) = Q(\mathbf{x})$ for all $\mathbf{x} \in \mathbb{R}^3$) is called $O(2, 1)$. The group of (not necessarily orientation preserving) isometries of H^+ is the index two subgroup of $O(2, 1)$ of maps which carry H^+ to itself, called the Lorentz group. To see why the Minkowski model agrees with the previous ones, one way is to compare the isometries, see eg [14].

The Klein model The Klein model (first however introduced by Beltrami!) is also a model in the disk, but now geodesics are Euclidean straight lines. It can be obtained from the hyperboloid model as follows. Any line in $\mathbb{R}^3$ through the origin O intersects H^+ in at most one point. The line also intersects the horizontal plane $x_0 = 1$ in one point. This defines a projection from H^+ to the unit disk. Since a geodesic in H^+ is the intersection of H^+ with a plane through O, the geodesic projects to a straight line in the Klein disk K. Thus geodesics in K are straight lines. Isometries of H^+ (elements of the Lorenz group) also induce isometries of K. This means that geodesics in K look straight, while on the other hand angles are distorted.

If K is placed as the equatorial disk inside the unit sphere in $\mathbb{R}^3$, then Euclidean orthogonal projection in the direction of the vertical vector $(1, 0, 0)$ sends K to the upper hemisphere, providing a map from K to the hemisphere model above.

0.4 Comparison between Euclidean, spherical and hyperbolic geometry

	Euclidean	Spherical	Hyperbolic
Basic givens	Points, lines, planes	”	”
Model	Euclidean plane	Surface of sphere	$\mathbb{D}$ or $\mathbb{H}$.
Lines	Euclidean lines	arcs of great circles	arcs orthogonal to boundary
Axioms. For example:	Any two distinct points lie on a unique line	”	”
Parallel lines	through any $P \notin L$ there is a unique line not meeting L	through $P \notin L$ there is no line not meeting L	through $P \notin L$ there are infinitely many lines not meeting L
Angle sum of triangle	π	$> \pi$	$< \pi$
Similar triangles	equal angles doesn't imply congruence	triangles with equal angles congruent	triangles with equal angles congruent
Circumference of circle	$2\pi r$	$2\pi \sin r$	$2\pi \sinh r$
Area of circle	πr^2	$4\pi \sin^2 r/2$	$4\pi \sinh^2 r/2$

0.5 A Potted History

Thales, c. 600 BC Started the formalised study of geometry.

Euclid, c. 300 BC The axiomatic formulation of geometry. ‘Givens’ were points, lines, straight lines, right angles etc.

Postulates:

1. Any two points lie on a unique line.
2. Any straight line can be continued indefinitely in either direction.
3. You can draw a circle of any centre and any radius.
4. All right angles are equal.
5. *The parallel postulate:* If a straight line, crossing another two straight lines L, L', makes angles α, β with L, L' on one side, and if $\alpha + \beta < \pi$, then L, L' if extended sufficiently far meet on that same side.

Attempts to prove the parallel postulate Starting from Greek times, many attempts were made to deduce the parallel postulate from the other axioms. It was shown to be equivalent to various other facts, for example:

1. Euclid. The angle sum of a triangle is π.
2. Proclus, 410-485 AD. Suppose L is a line and P is a point not on L. Then there exists a unique line L' through P and parallel to (ie not meeting) L. (Sometimes called *Playfair's axiom*, as in the 1795 edition of Euclid.)

   Exercise 0.2. *Show (i) and (ii) are equivalent to the parallel postulate.*
3. ibn al-Haytham, c. 1000 AD. The locus of points equidistant from a line is a line. (Discussed by Omar Khayyam, Iran, c. 1077)
4. Wallis, 1663. There exist similar triangles of different sizes.
5. Saccheri c.1733 and Lambert c.1766 explored many consequences of the *falsity* of the parallel postulate, in attempts to prove it by *reduction ad absurdum.*
6. Legendre (1752-1833) made what was probably the last attempt to prove it. Klügel (1763) expressed doubts that it could be proved.

The failure of the parallel postulate 1. By 1816, Gauss was expressing the conviction that the parallel postulate couldn't be proved. Gauss (c. 1822), Bolyai (1829) and Lobachevsky (1826) independently saw that they could develop a consistent geometry, so-called *non-Euclidean geometry*, in which the parallel postulate fails but the rest of Euclid's axioms remain true.

2. This led to a huge controversy, but gradually formulae and the structure of this new geometry were developed.
3. Beltrami (1868) settled the question by exhibiting (a) surfaces in $\mathbb{R}^3$ on which the geometry is 'non-Euclidean' and (b) models of the new geometry within Euclidean geometry (including the disk model in lectures). The conclusion is that if Euclidean geometry is consistent, then so is this new one.
4. Klein (1872) proposed a much broader view of geometry as the study of properties of a set invariant under a transformation group. He called the new geometry *hyperbolic* as its formulae can be obtained from those of spherical geometry by replacing trigonometric functions ($\cos, \tan$ etc) by the hyperbolic functions ($\cosh, \tanh$ etc.) Klein also entered the philosophical debate and partly through his work, the arguments about the 'existence' of non-Euclidean geometry were finally laid to rest.

Poincaré (c.1881) extended and refined Beltrami's model, realising that its isometries are just the linear fractional transformations from the disk $\mathbb{D}$ to itself. He also extended the model to 3 dimensions.

In 1904, Poincaré made the connection between symmetry groups of tessellations of the disk and the universal covering space of a surface with covering maps.

Hyperbolic geometry and surfaces Dehn (1912), Nielsen (1924), Koebe (1920's) developed Poincaré's ideas to use the existence of hyperbolic structures on surfaces to study the fundamental groups of surfaces and dynamics on surfaces. Hyperbolic geometry ideas were used for example by Artin (1924) prove existence of a dense geodesic on a surface, and by Hedlund and Hopf (1930's) to show ergodicity of geodesic flows. Connections with number theory and Diophantine approximation were also developed.

1920's-1950's Various special cases of hyperbolic 3-manifolds were studied.

1950's-1960's Ahlfors and Bers revived interest in the groups of isometries of 3-dimensional hyperbolic space because of its relation to the work of Teichmüller on complex (conformal) structures on surfaces in the 1940's.

Early 1970's Some remarkable examples of hyperbolic structures 3-manifolds were studied by Jørgensen and Riley. Marden systematically introduced the notions of 3-dimensional topology into the study of hyperbolic 3-manifolds.

1975-85 The Thurston revolution. Thurston showed that *very many* 3-manifolds have hyperbolic structures. His famous *geometrization conjecture* states that any 3-manifold can be cut, using topological data alone, into finitely many pieces, each of which carry one of eight special geometries, the most common of which is hyperbolic. Thurston introduced an enormous number of beautiful new ideas and vast developments followed.

2000-2006 Developments of Thurston's ideas resulted in the proofs of many old conjectures on hyperbolic 3-manifolds. Perleman's work led to a proof of Thurston's geometrization conjecture. The upshot is a virtually complete classification of geometric structures on 3-manifolds.

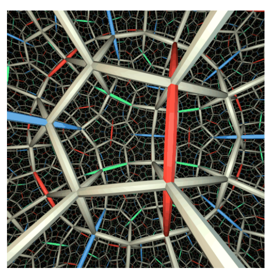

Figure 4: An internal veiw tiling of hyperbolic 3-space. Picture from the movie 'Not knot', courtesy Charlie Gunn.

Chapter 1

Linear Fractional Transformations

1.1 Linear fractional transformations

Good references for this chapter are [1], [2], [3] and [4].

Definition 1.1. *A* ***linear fractional transformation*** *or* ***Möbius map*** *is a map* $f : \hat{\mathbb{C}} = \mathbb{C} \cup \infty \longrightarrow \mathbb{C} \cup \infty$ *given by the formula* $f(z) = \dfrac{az+b}{cz+d}$ *where* $a, b, c, d \in \mathbb{C}$ *and* $ad - bc \neq 0$.

Here $\hat{\mathbb{C}}$ is the *Riemann sphere* identified with a sphere in $\mathbb{R}^3$ by stereographic projection, see Figure 2 in the Introduction. For more detailed discussion, see for example [3], [1] or [2].

Define

$$f(\infty) = \lim_{|z| \longrightarrow \infty} f(z) = \lim_{|z| \longrightarrow \infty} \frac{az+b}{cz+d} = \lim_{|z| \longrightarrow \infty} \frac{a + b/z}{c + d/z} = \frac{a}{c}.$$

Note also that $cz + d = 0 \iff z = -\dfrac{d}{c}$, in which case $az + b = -\dfrac{ad}{c} + b \neq 0$. So we can define $f(-\dfrac{d}{c}) = \infty$.

Remark 1.2. *To be rigourous, you have to replace* z *by* $1/w$ *in a neighbourhood of* ∞ *and to study the behaviour as* $w \longrightarrow 0$.

Such f is invertible, hence it is a bijection $\hat{\mathbb{C}} \longrightarrow \hat{\mathbb{C}}$. One easily calculates that the inverse map is given by

$$f^{-1}(z) = \frac{dz - b}{-cz + a}.$$

We can represent the coefficients of a Möbius map by the matrix $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{GL}(2, \mathbb{C})$. It is easy to check that you can compose Möbius maps by multiplying the matrices.

Observe that for any $\lambda \neq 0$, the maps

$$z \mapsto \frac{az + b}{cz + d} \quad \text{and} \quad z \mapsto \frac{\lambda az + \lambda b}{\lambda cz + \lambda d}$$

are the same function on $\hat{\mathbb{C}}$, while $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix} = \begin{pmatrix} \lambda a & \lambda b \\ \lambda c & \lambda d \end{pmatrix}$ with $\lambda \neq 0$. To fix the ambiguity, we can require that $ad - bc = 1$ which is achieved by dividing all coefficients by $\sqrt{\det A} = \sqrt{ad - bc}$. This still gives an ambiguity of multiplication by ± 1. Notice that to invert the map $f(z)$ above you don't need to divide all coefficients by $\sqrt{\det A} = \sqrt{ad - bc}$, as you would to invert A.

Definition 1.3. *A map $f : \hat{\mathbb{C}} \longrightarrow \hat{\mathbb{C}}$ is called **conformal** or **angle preserving** if, whenever smooth (C^1) curves $\mathcal{C}_1$ and $\mathcal{C}_2$ on $\hat{\mathbb{C}}$ meet at a point P with (signed) angle θ (measured from $\mathcal{C}_1$ to $\mathcal{C}_2$), then $f(\mathcal{C}_1)$ and $f(\mathcal{C}_2)$ meet at $f(P)$ with the same angle θ from $f(\mathcal{C}_1)$ to $f(\mathcal{C}_2)$.*

In particular, a conformal map is necessarily orientation preserving.

Definition 1.4. *We denote the set of all conformal bijections $\hat{\mathbb{C}} \longrightarrow \hat{\mathbb{C}}$ by $\mathrm{Aut}(\hat{\mathbb{C}})$.*

Proposition 1.5. *Möbius maps are conformal or angle preserving.*

Proof. This is true of any complex analytic map at any point z_0 at which $f'(z_0) \neq 0$. (Locally, the map expands by $|f'(z_0)|$ and rotates by $\operatorname{Arg} f'(z_0)$.) By computing, we see that $f'(z) = \dfrac{ad - bc}{(cz + d)^2} \neq 0$ except at $z = \infty$. By writing in terms of $w = 1/z$, you can also verify that f is conformal at ∞. $\square$

Exercise 1.6. *Complete the missing cases in the above proposition: Suppose that f is a Möbius map for which $f(\infty) \neq \infty$. Let $w = 1/z$ and $F(w) = f(1/z)$. Show that $F'(w) = -1/(c+dw)^2$ and hence that $F'(0) \neq 0$ unless $c = 0$. Deal with this last case by considering $G(w) = 1/F(w)$.*

Thus any Möbius map belongs to $\mathrm{Aut}(\hat{\mathbb{C}})$. Clearly $\mathrm{Aut}(\hat{\mathbb{C}})$ is a group under composition. The results above can be formalized as:

Proposition 1.7. *The map $F : \mathrm{GL}(2,\mathbb{C}) \longrightarrow \mathrm{Aut}(\hat{\mathbb{C}})$ defined by*

$$F : \begin{pmatrix} a & b \\ c & d \end{pmatrix} \mapsto \left(f(z) = \frac{az+b}{cz+d} \right)$$

is a homomorphism with kernel

$$\mathrm{Ker}\, F = \left\{ \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix} : \lambda \in \mathbb{C}^* \right\}.$$

Proof. Saying that F is an homomorphism just means that we can compose Möbius maps by multiplying matrices. (Exercise: Check this.)

To find the kernel:

$$\frac{az+b}{cz+d} \equiv z \iff cz^2 + (d-a)z + b \equiv 0 \iff c = b = 0, d = a.$$

□

A much deeper fact is that the map F is *surjective*, that is, that *every* conformal bijection of $\hat{\mathbb{C}}$ can be represented by a Möbius map:

Theorem 1.8. *The map $F : \mathrm{GL}(2,\mathbb{C}) \longrightarrow \mathrm{Aut}(\hat{\mathbb{C}})$ defined above is a surjection.*

Proof. This requires Liouville's theorem and other results from complex analysis, see Example sheet 1, [3] (p. 9) or [4] (p. 5). □

Usually, to avoid the ambiguities in of representation by matrices, we normalise Möbius maps so their coefficients are in

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

As above, this can be achieved by dividing all coefficients by $\sqrt{\det(A)}$. This still leaves the ambiguity of multiplication by ± 1. Defining $\mathrm{PSL}(2,\mathbb{C}) = \mathrm{SL}(2,\mathbb{C})/\pm I$ we have:

Corollary 1.9 (Normalisation)**.** *(1) Any $T \in \mathrm{Aut}(\hat{\mathbb{C}})$ can be represented by a matrix* $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}(2, \mathbb{C})$.
(2) $\mathrm{Aut}(\hat{\mathbb{C}}) \cong \mathrm{PSL}(2, \mathbb{C})$.

Proof. (1) By Theorem 1.8, any $T \in \mathrm{Aut}(\hat{\mathbb{C}})$ can be represented by $M = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{GL}(2, \mathbb{C})$. Let $\Delta = \det(M) = ad - bc$. The matrices M and $\frac{1}{\sqrt{\Delta}}M = \begin{pmatrix} \frac{a}{\sqrt{\Delta}} & \frac{b}{\sqrt{\Delta}} \\ \frac{c}{\sqrt{\Delta}} & \frac{d}{\sqrt{\Delta}} \end{pmatrix}$ have the same image under F in $\mathrm{Aut}(\hat{\mathbb{C}})$, and $\det(\frac{M}{\sqrt{\Delta}}) = 1$.
(2) The obvious map $\mathrm{SL}(2, \mathbb{C}) \longrightarrow \mathrm{Aut}(\hat{\mathbb{C}})$ has kernel $\mathrm{SL}(2, \mathbb{C}) \cap \mathrm{Ker} F = \pm I$. □

1.1.1 Cross-ratio and transitivity of $\mathrm{Aut}(\hat{\mathbb{C}})$

The group $\mathrm{Aut}(\hat{\mathbb{C}})$ includes some especially simple transformations:
(1) *translations* $z \mapsto z + t$ where $t \in \mathbb{C}$;
(2) *rotations* around $a \in \mathbb{C}$: $z \mapsto e^{i\theta}(z - a) + a$;
(3) *similarities* $z \mapsto \lambda z$ with $\lambda > 0$. (This is an expansion if $\lambda > 1$ and a contraction if $\lambda < 10$.)

These transformations all fix ∞. It also includes:
(4) *inversion* $z \mapsto 1/z$ which interchanges 0 and ∞.

Exercise 1.10. *By writing*

$$T(z) = \frac{az + b}{cz + d} = \frac{a}{c} - \frac{1}{c(cz + d)},$$

show that any $T \in \mathrm{Aut}(\hat{\mathbb{C}})$ is a composition of the above four types.

Definition 1.11. *A group G is said to act **transitively** on a space X if, given two points x_0 and x_1 in X, there exists $g \in G$ such that $g(x_0) = x_1$. It acts **freely** if for all $x_0 \in X$, $g(x_0) = x_0$ implies $g = \mathrm{id}$ for all $x_0 \in X$.*

Exercise 1.12. *Use the above transformations to show that $\mathrm{Aut}(\hat{\mathbb{C}})$ acts transitively but not freely on $\hat{\mathbb{C}}$. Now do the same for ordered pairs of distinct points $(z_1, z_2) \in \hat{\mathbb{C}}$.*

Proposition 1.13. $\mathrm{Aut}(\hat{\mathbb{C}})$ *acts transitively and freely on ordered triples of distinct points in* $\hat{\mathbb{C}}$. *More precisely, given two ordered sets of distinct points* (z_2, z_3, z_4) *and* (w_2, w_3, w_4) *in* $\hat{\mathbb{C}}^3$, *there exists a unique* $T \in \mathrm{Aut}(\hat{\mathbb{C}})$ *such that* $T(z_i) = w_i$ *for* $i = 2, 3, 4$.

Proof. • *Existence*: First, we prove existence in the case $(w_2, w_3, w_4) = (0, \infty, 1)$. Denote by $T_{(z_2,z_3,z_4)}$ the transformation $T(z) = \frac{z - z_2}{z - z_3} \cdot \frac{z_4 - z_3}{z_4 - z_2}$. Check that $T_{(z_2,z_3,z_4)}$ maps $(z_2, z_3, z_4) \mapsto (0, \infty, 1)$. In the general case the transformation T we need is obtained by composing $T_{(z_2,z_3,z_4)}$ and $T^{-1}_{(w_2,w_3,w_4)}$, that is $T = T^{-1}_{(w_2,w_3,w_4)} \circ T_{(z_2,z_3,z_4)}$.

• *Uniqueness*: First we prove that, given $S \in \mathrm{Aut}(\hat{\mathbb{C}})$ such that $S(0) = 0, S(\infty) = \infty$ and $S(1) = 1$, then $S = id$. Let $S(z) = \frac{az + b}{cz + d}$. If $S(0) = b/d = 0$ and $S(\infty) = a/c = \infty$, then $b = c = 0$. If $S(1) = \frac{a + b}{c + d} = \frac{a}{d} = 1$, then $a = d$. Since we may assume that $\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = 1$, this gives $ad = 1$, so $S = id$.

Now suppose there were two transformations T, T' such that $(z_2, z_3, z_4) \mapsto (w_2, w_3, w_4)$. Then $T_{(w_2,w_3,w_4)}TT^{-1}_{(z_2,z_3,z_4)}, T_{(w_2,w_3,w_4)}T'T^{-1}_{(z_2,z_3,z_4)}$ both fix $(0, \infty, 1)$ and the result follows from the above. □

Definition 1.14. *Let* z_1, z_2, z_3, z_4 *be an ordered set of distinct points in* $\hat{\mathbb{C}}$. *The* ***cross-ratio*** *of* z_1, z_2, z_3, z_4 *is defined by*

$$[z_1, z_2, z_3, z_4] = \frac{z_1 - z_2}{z_1 - z_3} \cdot \frac{z_4 - z_3}{z_4 - z_2} \in \mathbb{C}.$$

(Why is $[z_1, z_2, z_3, z_4] \neq \infty$ for any z_1, z_2, z_3, z_4?) The map $T_{(z_2,z_3,z_4)}$ used in the proof of Proposition 1.13 can be nicely expressed using the cross-ratio:

$$T_{(z_2,z_3,z_4)}(z) = \frac{z - z_2}{z - z_3} \cdot \frac{z_4 - z_3}{z_4 - z_2} = [z, z_2, z_3, z_4].$$

Corollary 1.15. *The cross-ratio is invariant under* $\mathrm{Aut}(\hat{\mathbb{C}})$, *that is if* $T \in \mathrm{Aut}(\hat{\mathbb{C}})$, *then* $[z_1, z_2, z_3, z_4] = [T(z_1), T(z_2), T(z_3), T(z_4)]$.

Proof. Method 1: Let $T(z) = \frac{az + b}{cz + d}$. The result follows by a direct calculation substituting in the two sides.

Method 2: Define $S \in \mathrm{Aut}(\hat{\mathbb{C}})$ by the formula $S(z) = [z, z_2, z_3, z_4]$. Then S sends $z_2 \mapsto 0, z_3 \mapsto \infty, z_4 \mapsto 1$ and by Proposition 1.13, S is the unique element of $\mathrm{Aut}(\hat{\mathbb{C}})$ with this property.

Now write $w_i = T(z_i)$ for $i = 2, 3, 4$ and let $W(z) = [z, w_2, w_3, w_4]$. So W sends $w_2 \mapsto 0, w_3 \mapsto \infty, w_4 \mapsto 1$, hence $W \circ T$ sends $z_2 \mapsto 0, z_3 \mapsto \infty, z_4 \mapsto 1$, so $W \circ T = S$. Now $W \circ T(z_1) = S(z_1) = [z_1, z_2, z_3, z_4]$ while on the other hand, $W \circ T(z_1) = [T(z_1), w_2, w_3, w_4] = [T(z_1), T(z_2), T(z_3), T(z_4)]$. □

Remark 1.16. This crucial property of invariance of cross-ratio will allow us to define hyperbolic distance. Notice that the Euclidean distance between 2 points is invariant under translations or rotations and that the ratio of the distances between 3 points is invariant under Euclidean similarities.

Lemma 1.17. *Four distinct points $z_1, z_2, z_3, z_4 \in \hat{\mathbb{C}}$ are concyclic if and only if $[z_1, z_2, z_3, z_4] \in \mathbb{R}$.*

Remark 1.18. *(i) Here "concyclic" includes "collinear", since we can view a line as a circle through $\infty \in \hat{\mathbb{C}}$. (Lines through ∞ are exactly the stereographic projections of great circles through the north pole.)*
(ii) Permuting the order of the points $z_1, z_2, z_3, z_4 \in \hat{\mathbb{C}}$ may change $[z_1, z_2, z_3, z_4]$, but it will not change the fact that $[z_1, z_2, z_3, z_4] \in \mathbb{R}$, see Exercise 1.20.

Proof of Lemma 1.17. Suppose that $z_1, z_2, z_3, z_4 \in \hat{\mathbb{C}}$ are concyclic, then (up to reflection) they must be arranged in one or other of the configurations in Figure 1.1, in which the angles are found by using elementary geometry in circles. For the configuration shown on the left: $\mathrm{Arg}\Big(\dfrac{z_1 - z_2}{z_1 - z_3}\Big) = -\theta = \mathrm{Arg}\Big(\dfrac{z_4 - z_2}{z_4 - z_3}\Big)$. Hence

$$\mathrm{Arg}\Big(\frac{z_1 - z_2}{z_1 - z_3} \cdot \frac{z_4 - z_3}{z_4 - z_2}\Big) = -\theta + \theta = 0 \Rightarrow [z_1, z_2, z_3, z_4] > 0.$$

For the configuration on the right: $\mathrm{Arg}\Big(\dfrac{z_1 - z_2}{z_1 - z_3}\Big) = \theta$ and $\mathrm{Arg}(\dfrac{z_4 - z_2}{z_4 - z_3}\Big) = -(\pi - \theta)$. Hence

$$\mathrm{Arg}\Big(\frac{z_1 - z_2}{z_1 - z_3} \cdot \frac{z_4 - z_3}{z_4 - z_2}\Big) = \theta + (\pi - \theta) = \pi$$

from which $[z_1, z_2, z_3, z_4] < 0$.

The converse is similar. □

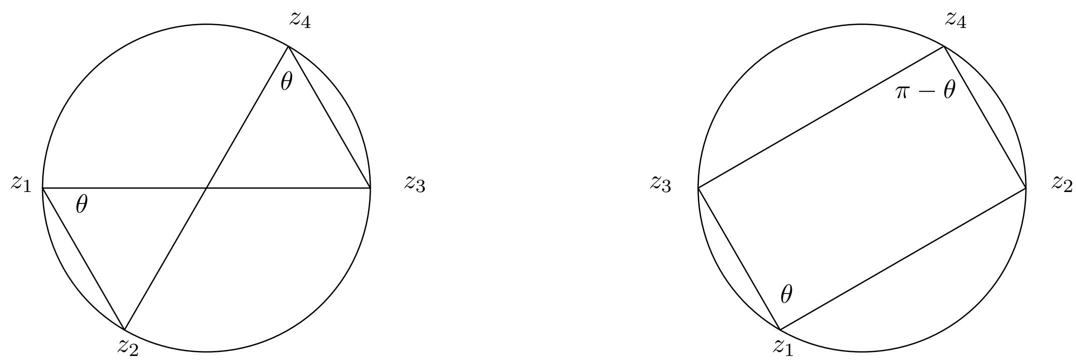

Figure 1.1: Concyclic points have real cross ratio.

Corollary 1.19. *Suppose $T \in \mathrm{Aut}(\hat{\mathbb{C}})$. The T maps circles and lines to circles and lines.*

Proof. Method 1: The result follows immediately from Corollary 1.15 and Lemma 1.17.
Method 2: By direct calculation. The equation of a circle centre a radius r is $(z-a)(\bar{z}-\bar{a}) = r^2$. Check that this can be rearranged as $|z|^2+Bz+\bar{B}\bar{z}+C = 0$ where $|B|^2 - C > 0$, and that conversely any equation in this second form is a circle. Likewise show that a line can be written $\lambda(z - a) = \overline{\lambda(z - a)}$ or equivalently $Bz + \bar{B}\bar{z} + C = 0$. Using Exercise 1.10, show you have only to check the calculation for the transformation $z \mapsto \dfrac{1}{z}$, which is easy. □

Exercise 1.20 (Effect of permutation on the cross-ratio). *Check that if $\chi = [z_1, z_2, z_3, z_4]$, then by permuting the z_i we obtain exactly the six values $\chi, \frac{1}{\chi}, 1 - \chi, \frac{1}{1-\chi}, \frac{\chi}{\chi-1}, \frac{\chi-1}{\chi}$.*

Exercise 1.21 (Orientation reversing (anticonformal) automorphisms of $\hat{\mathbb{C}}$). *One such map is $z \mapsto \bar{z}$. Check that the most general such map is $z \mapsto \dfrac{a\bar{z} + b}{c\bar{z} + d}$. Prove that any such map takes $[z_1, z_2, z_3, z_4]$ to $\overline{[z_1, z_2, z_3, z_4]}$ and maps circles to circles.*

1.2 Automorphisms of the unit disk $\mathbb{D}$

Let $\mathbb{D}$ denote the unit disk $\{z \in \mathbb{C} : |z| < 1\}$ and set

$$\mathrm{Aut}(\mathbb{D}) = \left\{ T \in \mathrm{Aut}(\hat{\mathbb{C}}) : T(\mathbb{D}) = \mathbb{D} \right\}.$$

Notice that any $T \in \mathrm{Aut}(\hat{\mathbb{C}})$ is automatically a bijection from $\mathbb{D}$ to itself. We are interested in $\mathrm{Aut}(\mathbb{D})$ because, as we shall see in the next chapter, it can be viewed as the set of orientation preserving rigid motions of the Poincaré disk model of hyperbolic space $\mathbb{D}$ with the hyperbolic metric.

Theorem 1.22. *The set* $\mathrm{Aut}(\mathbb{D})$ *is the subgroup of* $\mathrm{Aut}(\hat{\mathbb{C}})$ *of Möbius maps of the form*

$$f(z) = \frac{e^{i\theta}(z-a)}{1-\bar{a}z} \quad \text{with} \quad e^{i\theta} \in \mathbb{S}^1, a \in \mathbb{D}, \tag{1.1}$$

or equivalently of the form

$$f(z) = \frac{az+b}{\bar{b}z+\bar{a}} \quad \text{with} \quad |a|^2 - |b|^2 = 1. \tag{1.2}$$

Exercise 1.23. *Check that functions of type* (1.1) *form a group. Also explain why the forms* (1.1) *and* (1.2) *are equivalent.*

Remark 1.24. *The set*

$$\left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{GL}(2,\mathbb{C}) : d = \bar{a}, c = \bar{d}, |a|^2 - |b|^2 = 1 \right\}$$

is called $\mathrm{SU}(1,1)$.

Proof. Suppose that f is of the form (1.1). We check:

$$|f(z)| < 1 \iff (z-a)(\bar{z}-\bar{a}) < (1-\bar{a}z)(1-a\bar{z}) \iff (1-z\bar{z})(1-a\bar{a}) > 0.$$

So if $|a| < 1$, then $|f(z)| < 1 \iff |z| < 1$. Hence f restricts to a bijective map $\mathbb{D} \longrightarrow \mathbb{D}$, that is $f \in \mathrm{Aut}(\mathbb{D})$.

Conversely, to show that any conformal automorphism $g : \mathbb{D} \longrightarrow \mathbb{D}$ is of this form, first let $f(z) = \dfrac{(z-a)}{1-\bar{a}z}$ where $a = g^{-1}(0)$. Then $f(a) = 0$, so

$g \circ f^{-1}(0) = 0$ and $g \circ f^{-1}$ is a conformal automorphism. By the Schwarz Lemma from Complex Analysis, if $h = g \circ f^{-1} : \mathbb{D} \longrightarrow \mathbb{D}$ and $h(0) = 0$, then $|h(z)| \leqslant |z|$ and $|z| = |h^{-1} \circ h(z)| \leqslant |h(z)|$. So $|h(z)| = |z|$. From Schwarz' Lemma again, it follows that $h(z) = e^{i\theta} z$ for some $\theta \in [0, 2\pi)$. Thus $g \circ f^{-1}(z) = e^{i\theta} z$ and the result follows. □

1.3 Automorphisms of the upper half plane

The upper half plane $\mathbb{H}$ is by definition $\{z \in \mathbb{C} : \Im(z) > 0\}$. As we shall see in Chapter 2, $\mathrm{Aut}(\mathbb{H}) = \left\{T \in \mathrm{Aut}(\hat{\mathbb{C}}) : T(\mathbb{H}) = \mathbb{H}\right\}$ is the set of orientation preserving rigid motions of the upper half plane model of hyperbolic space.

To move from $\mathbb{D}$ to $\mathbb{H}$ we introduce the *Cayley transformation*

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

Lemma 1.25. *The Cayley transformation induces a conformal automorphism* $\mathbb{H} \longrightarrow \mathbb{D}$*.*

Proof. We check easily that $C(\infty) = 1, C(1) = -i, C(0) = -1$. It follows that C maps the line $\mathbb{R} \cup \infty$ to the circle through $1, -i, -1$, that is, to $\partial\mathbb{D}$. Thus C must map the two connected components of $\hat{\mathbb{C}} \setminus \mathbb{R} \cup \infty$ to the two connected components of $\hat{\mathbb{C}} \setminus \partial\mathbb{D}$. To see which goes to which, note that $C(i) = 0$ so $\mathbb{H}$ must map to $\mathbb{D}$ as claimed. □

Exercise 1.26. *Check directly that if* $z = u + iv \in \mathbb{H}$ *and* $w = C(z)$ *then* $w \in \mathbb{D}$*, and conversely.*

Theorem 1.27. *The set* $\mathrm{Aut}(\mathbb{H})$ *of analytic bijections* $\mathbb{H} \longrightarrow \mathbb{H}$ *is the subgroup of* $\mathrm{Aut}(\hat{\mathbb{C}})$ *of Möbius maps of the form:*

$$f(z) = \frac{az + b}{cz + d} \quad \text{with} \quad a, b, c, d \in \mathbb{R}, ad - bc > 0. \tag{1.3}$$

Normalising, this shows that $\mathrm{Aut}(\mathbb{H}) = \mathrm{SL}(2, \mathbb{R}) / \pm I = \mathrm{PSL}(2, \mathbb{R})$.

Proof. First we first check that any map f of the form (1.3) is a bijection of $\mathbb{H}$ to $\mathbb{H}$. We compute that $\Im f(z) = \dfrac{\Im z(ad - bc)}{(cz + d)^2}$. Since by assumption $ad - bc > 0$ this means that $f(\mathbb{H}) \subset \mathbb{H}$. Now since f is a bijection from $\hat{\mathbb{C}}$ to itself, to see that it is also a bijection of $\mathbb{H}$ to $\mathbb{H}$, we just have to check it is surjective. This follows since by the same reasoning as before, $f^{-1}(\mathbb{H}) \subset \mathbb{H}$. So $f \in \mathrm{Aut}(\mathbb{H})$.

Now let us prove the final statement. Since $ad - bc > 0$, we have $\sqrt{ad-bc} \in \mathbb{R}$. So if we normalise the matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{GL}(2,\mathbb{R})$ by dividing each coefficient by $\sqrt{ad-bc}$, we see that any map f of the form (1.27) is the image of under F (see Theorem 1.8) of a map in $\mathrm{SL}(2,\mathbb{R})$. So $\mathrm{Aut}(\mathbb{H}) \supset \mathrm{SL}(2,\mathbb{R})/\pm I = \mathrm{PSL}(2,\mathbb{R})$.

Finally we have to show that an arbitrary $F \in \mathrm{Aut}(\mathbb{H})$ is of the above form. Defining $h : z \mapsto \dfrac{z - \Re F(i)}{\Im F(i)}$ we find $h \circ F \in \mathrm{Aut}(\mathbb{H})$ and $h \circ F(i) = i$. (Why is $h \in \mathrm{Aut}(\mathbb{H})$?)

Now composing with a rotation $g : z \mapsto \dfrac{\cos(\theta)z + \sin(\theta)}{-\sin(\theta)z + \cos(\theta)}$, you obtain a function $T = g \circ h \circ F$ such that $T(i) = i$ and $T'(i) = 1$. Then check that $\hat{T} = C \circ T \circ C^{-1} \in \mathrm{Aut}(\mathbb{D}), \hat{T}(0) = 0, \hat{T}'(0) = 1$. Using form (1.1) of Theorem 1.22, that this implies that $\hat{T} = id$ and so $T = id$. Thus $F = h^{-1} \circ g^{-1}$ and you can now easily verify that F is of the form (1.3). □

Remark 1.28. The above theorem can also be proved by conjugating to $\mathbb{D}$ using the Cayley transform. The main point is to verify that any $T \in \mathrm{Aut}(\mathbb{D})$ conjugates to some element in $\mathrm{SL}(2,\mathbb{R})$. Write

$$T = \begin{pmatrix} a & b \\ \bar{b} & \bar{a} \end{pmatrix} \text{ with } \ |a|^2 - |b|^2 = 1.$$

As a matrix the Cayley transform $C : \mathbb{H} \longrightarrow \mathbb{D}$ is written

$$C = \begin{pmatrix} 1/\sqrt{\Delta} & -i/\sqrt{\Delta} \\ 1/\sqrt{\Delta} & i/\sqrt{\Delta} \end{pmatrix}$$

where $\Delta = 2i$ is the normalising factor which gives $\det C = 1$. Calculating we find eventually that

$$C^{-1}TC = \begin{pmatrix} \Re a + \Re b & \Im a - \Im b \\ -\Im a - \Im b & \Re a - \Re b \end{pmatrix}$$

which one easily sees (using $|a|^2 - |b|^2 = 1$) is in $\mathrm{SL}(2, \mathbb{R})$.

Now given any $S \in \mathrm{Aut}(\mathbb{H})$ we have $CSC^{-1} \in \mathrm{Aut}(\mathbb{D})$ so that CSC^{-1} can be equated with some T as above. We have just calculated that $C^{-1}TC = C^{-1}CSC^{-1}C = S \in \mathrm{SL}(2, \mathbb{R})$ and we are done.

Chapter 2

Basics of hyperbolic geometry

2.1 Metrics, geodesics and isometries.

We shall study the upper half plane and unit disk models of hyperbolic space at the same time. The models are:

(i) $\mathbb{H} = \{z \in \mathbb{C} : \Im(z) > 0\}$ with the metric $ds^2 = \dfrac{dx^2 + dy^2}{y^2}$;

(ii) $\mathbb{D} = \{z \in \mathbb{C} : |z| < 1\}$ with the metric $ds^2 = \dfrac{4(dx^2 + dy^2)}{(1 - (x^2 + y^2))^2}$.

Proposition 2.1. *With the above metrics, both* $\mathbb{H}$ *and* $\mathbb{D}$ *are spaces of constant negative curvature* -1.

Proof. We compute using basic formulae from differential geometry. If $ds^2 = g_1 dx^2 + g_2 dy^2$ is a Riemannian metric on a surface, then the curvature K is given by the formula:

$$K = \frac{-1}{\sqrt{g_1 g_2}} \left[\frac{\partial}{\partial x} \left(\frac{1}{\sqrt{g_1}} \frac{\partial}{\partial x} \sqrt{g_2} \right) + \frac{\partial}{\partial y} \left(\frac{1}{\sqrt{g_2}} \frac{\partial}{\partial y} \sqrt{g_1} \right) \right].$$

Applying this to our metric on $\mathbb{H}$ we have $g_1 = g_2 = g = 1/y^2$, so that

$$\frac{\partial}{\partial x}\sqrt{g} = 0, \quad \frac{\partial}{\partial y}\sqrt{g} = -\frac{1}{y^2}, \quad \frac{1}{\sqrt{g}}\frac{\partial}{\partial y}\sqrt{g} = -\frac{1}{y} \quad \frac{\partial}{\partial y}\left(\frac{1}{\sqrt{g}}\frac{\partial}{\partial y}\sqrt{g}\right) = \frac{1}{y^2}$$

and finally

$$K = -\frac{1}{g} \cdot \frac{\partial}{\partial y}\left(\frac{1}{\sqrt{g}}\frac{\partial}{\partial y}\sqrt{g}\right) = -1.$$

Exercise: Make a similar computation in $\mathbb{D}$. $\square$

The expression for ds gives us the length of an arc γ in $\mathbb{H}$ by the formula $l(\gamma) = \int_\gamma ds$. More precisely, if $\gamma : [t_0, t_1] \longrightarrow \mathbb{H}$ with $\gamma(t) = x(t) + iy(t) \in \mathbb{H}$, then

$$l(\gamma) := \int_{t_0}^{t_1} \frac{1}{y(t)} \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2} dt.$$

For more details about this, see books on differential geometry or [6] Ch. 3. We shall be doing some explicit examples below.

2.1.1 Geodesics and isometries

The distance $d_{\mathbb{D}}$ or $d_{\mathbb{H}}$ between two points P, Q is defined to be $\inf l(\gamma)$ where the infimum is taken over all paths γ joining P to Q.

Definition 2.2. *A path γ from z_0 to z_1 is a **geodesic** if it locally minimises distances.*

Notice that in general a path may locally minimise distance without actually giving the minimum. For example, there are in general two geodesics joining any two points on a sphere (corresponding to the two different ways of going round the great circle from one point to the other). Both routes are geodesic but unless the points are diametrically opposite, one is shorter than the other.

As we shall see, in hyperbolic space there is only one geodesic arc between any two points, so geodesics do automatically minimise distance. First we prove:

Proposition 2.3. *(i) Vertical lines are geodesics in $\mathbb{H}$. Moreover if $b > a$, then $d_{\mathbb{H}}(ai, bi) = \log b/a$.*

(ii) Radial lines are geodesics in $\mathbb{D}$. In particular $d_{\mathbb{D}}(0, a) = \log\left(\frac{1+a}{1-a}\right)$.

Proof. (i) Consider any path γ joining ai and bi on the imaginary axis $i\mathbb{R}$. (The proof is the same for any other vertical line.) Then

$$l(\gamma) = \int_{y=a}^{y=b} \frac{1}{y(t)} \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2} dt \geqslant \int_a^b \frac{1}{y(t)} \sqrt{\left(\frac{dy}{dt}\right)^2} dt$$

$$= \int_a^b \frac{1}{y} \frac{dy}{dt} dt = \int_a^b \frac{dy}{y} = [\log y]_a^b = \log\left(\frac{b}{a}\right)$$

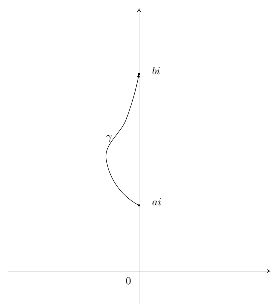

Figure 2.1: An arc γ joining two points on the imaginary axis in $\mathbb{H}$.

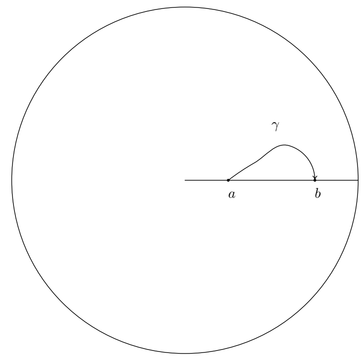

Figure 2.2: An arc γ joining two points on the horizontal radius in $\mathbb{D}$.

with equality if and only if γ is the vertical line (Why?). This argument shows that the vertical line from ia to ib actually minimises the distance, both locally and globally, so we deduce that

$$d_{\mathbb{H}}(ai, bi) = \log b/a.$$

(ii) Take polar coordinates (r, θ) in $\mathbb{D}$ and recall that $dx^2 + dy^2 = dr^2 + r^2 d\theta^2$. Then

$$l(\gamma) = \int_a^b \frac{2}{1-r^2} \sqrt{\left(\frac{dr}{dt}\right)^2 + r\left(\frac{d\theta}{dt}\right)^2}\, dt \geqslant \int_a^b \frac{2}{1-r^2} dr$$

$$= \int_a^b \left(\frac{1}{1-r} + \frac{1}{1+r}\right) dr = \left[\log\left(\frac{1+r}{1-r}\right)\right]_a^b.$$

In particular we deduce as in (i) that $d_{\mathbb{D}}(0, a) = \log\left(\frac{1+a}{1-a}\right)$. □

The boundary at infinity

It follows from Proposition 2.3 that for any point $x + i \in \mathbb{H}$ on the same horizontal level as i, we have $d_{\mathbb{H}}(x + i, x + it) = -\log t \longrightarrow \infty$ as $t \longrightarrow 0$. In other words, the distance from $x + i$ to the point $x \in \mathbb{R}$ vertically below it is infinite. For this reason, $\hat{\mathbb{R}} = \mathbb{R} \cup \infty$ is called the *boundary at infinity* of $\mathbb{H}$,

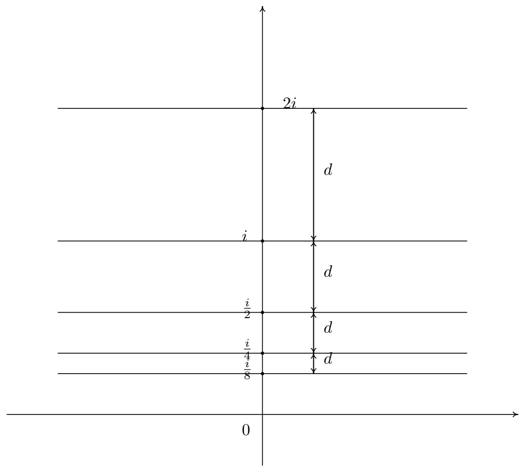

Figure 2.3: Equally spaced horizontal lines in $\mathbb{H}$ appear to get exponentially closer as we approach the real axis.

written $\partial\mathbb{H}$. Point on $\partial\mathbb{H}$ are sometimes called *ideal points*. Figure 2.3, which shows equally spaced horizontal levels in $\mathbb{H}$, illustrates the infinite distance from a point in $\mathbb{H}$ to the boundary.

In the disk $\mathbb{D}$, set $R = d_{\mathbb{D}}(0, r)$. From the formula in Proposition 2.3 we get $e^R = \dfrac{1+r}{1-r}$. Then

$$r = \frac{e^R - 1}{e^R + 1} = \tanh \frac{R}{2}. \tag{2.1}$$

Thus

$$R \longrightarrow \infty \iff r \longrightarrow 1$$

and we similarly call $\partial\mathbb{D}$ the *boundary at infinity* of $\mathbb{D}$.

Exercise 2.4. *Prove that concentric circles centre $O \in \mathbb{D}$ which are at equal hyperbolic distance apart become exponentially clese together in the Euclidean metric.*

Isometries

Before we find the other geodesics in $\mathbb{H}$ and $\mathbb{D}$, we first investigate isometries in the two models. Formally, a map $f : \mathbb{H} \longrightarrow \mathbb{H}$ is an *orientation preserving* *isometry* of $\mathbb{H}$ if it is differentiable (as a function $\mathbb{R}^2 \longrightarrow \mathbb{R}^2$), if $\det Df(z) > 0 \; \forall z \in \mathbb{H}$, and if

$$d(f(z_1), f(z_2)) = d(z_1, z_2) \text{ for all } \; z_1, z_2 \in \mathbb{H}.$$

The set of all orientation preserving isometries of $\mathbb{H}$ is denoted $\mathrm{Isom}^+(\mathbb{H})$. The set of orientation preserving isometries of $\mathbb{D}$, $\mathrm{Isom}^+(\mathbb{D})$, is defined similarly.

Recall from Chapter 1 that $\mathrm{Aut}(\mathbb{H}) \cong \mathrm{PSL}(2, \mathbb{R})$ and $\mathrm{Aut}(\mathbb{D}) \cong \mathrm{PSU}(1,1)$ are the subgroups of linear fractional transformations which map $\mathbb{H}$ and $\mathbb{D}$ to themselves respectively.

Proposition 2.5. *The groups* $\mathrm{Aut}(\mathbb{H})$ *and* $\mathrm{Aut}(\mathbb{D})$ *act by isometries on* $\mathbb{H}, \mathbb{D}$ *respectively.*

Proof. (i) By Theorem 1.27, if $T \in \mathrm{Aut}(\mathbb{H})$ then $T(z) = w = \dfrac{az+b}{cz+d}$ with $a, b, c, d \in \mathbb{R}, ad - bc = 1$. Recall that $\Im w = \dfrac{\Im z}{|cz+d|^2}$. (This means that if $z \in \mathbb{H}$, then $T(z) \in \mathbb{H}$.) Now we also compute that $dw = \dfrac{dz}{(cz+d)^2}$ and apply the change of variable formula. Let $T : \mathbb{H} \longrightarrow \mathbb{H}$ defined by $T(z) = w$. So

$$l\left(T(\gamma)\right) = \int_{T(\gamma)} \frac{|dw|}{\Im w} = \int_\gamma \frac{|cz+d|^2}{\Im z} \cdot \frac{|dz|}{|cz+d|^2} = l(\gamma).$$

See [6] or basic differential geometry books for more details.

(ii) By Theorem 1.22, if $T \in \mathrm{Aut}(\mathbb{D})$ then $T(z) = w = \dfrac{e^{i\theta}(z-a)}{1-\bar{a}z}$ with $e^{i\theta} \in \mathbb{S}^1$ and $a \in \mathbb{D}$. Then

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

and

$$1 - |w|^2 = 1 - \left|\frac{(z-a)}{1-\bar{a}z}\right|^2 = \frac{(1-|z|^2)(1-|a|^2)}{|1-\bar{a}z|^2}.$$

So

$$\frac{dw}{1-|w|^2} = \frac{dz}{1-|z|^2}.$$

Now you can use the change of variable formula as before. □

After we have discussed hyperbolic circles, we shall prove in Theorem 2.11 that in fact

$$\mathrm{Isom}^+(\mathbb{H}) \cong \mathrm{PSL}(2,\mathbb{R}) \text{ and } \quad \mathrm{Isom}^+(\mathbb{D}) \cong \mathrm{PSU}(1,1).$$

To avoid proving many different facts twice, as in the above theorem, we can use the *Cayley transform* which we introduced at the end of Chapter 1.

Proposition 2.6. *The Cayley transformation $C : \mathbb{H} \longrightarrow \mathbb{D}$, defined by*

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

is an isometry from $\mathbb{H} \longrightarrow \mathbb{D}$.

Proof. We already showed in Lemma 1.25 that C is a bijection between $\mathbb{H}$ and $\mathbb{D}$. Let $z = C^{-1}(w) = \dfrac{i(w+1)}{1-w}$. Then

$$\frac{dz}{dw} = \frac{2i}{(1-w)^2}, \quad \Im z = \Im\left[\frac{i(w+1)}{1-w}\right] = \Im\left[\frac{i(w+1)(1-\bar{w})}{|1-w|^2}\right] = \frac{1-|w|^2}{|1-w|^2}.$$

So

$$\frac{|dz|}{\Im z} = \frac{2|dw|}{|1-w|^2} \cdot \frac{|1-w|^2}{1-|w|^2} = \frac{2|dw|}{1-|w|^2}.$$

The result follows using the change of variable formula as in the proof of Proposition 2.5. □

Exercise 2.7. *Explain why Proposition 2.6 gives an alternative proof of Proposition 2.5 (ii) above.*

Geodesics

Now we are in a position to find all the other geodesics besides the vertical lines and radii we discussed above. Geodesics are described in the following proposition, illustrated in Figure 2.4.

Proposition 2.8. *(i) Let $P, Q \in \mathbb{H}$. If $\Re P = \Re Q$ then the geodesic from P to Q is the vertical line from P to Q, otherwise it is the arc of circle, with centre on $\mathbb{R} \cup \infty = \partial\mathbb{H}$, joining P to Q.*
(ii) Let $P, Q \in \mathbb{D}$. If P, Q are on the same diameter then the geodesic from P to Q is the Euclidean line segment joining them, otherwise it is the arc of circle, orthogonal to $\mathbb{S}^1$, joining P to Q.
In all cases, the geodesic from P to Q is unique.

Proof. The first thing to notice is that an isometry carries a geodesic to a geodesic, so throughout the proof we use isometries to move objects into a convenient position.

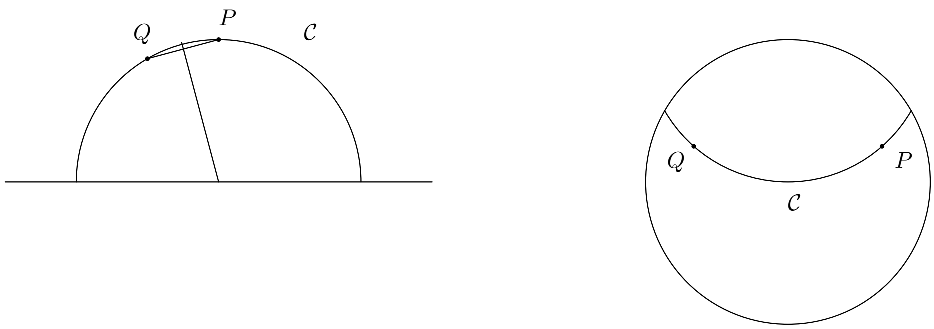

Figure 2.4: Geodesics joining points P, Q in $\mathbb{H}$ and $\mathbb{D}$.

(i) Let $\eta < \xi$ be the endpoints of the semicircle $\mathcal{C}$, with centre on $\mathbb{R} = \partial\mathbb{H}$, joining P to Q (as described in the figure) and let $T : z \mapsto \dfrac{z-\xi}{z-\eta}$. Then $T \in \mathrm{Aut}(\mathbb{H})$ because $\det \begin{pmatrix} 1 & -\xi \\ 1 & -\eta \end{pmatrix} = \xi - \eta > 0$. Moreover T carries the semicircle $\mathcal{C}$ to the imaginary axis because $T(\xi) = 0, T(\eta) = \infty$ and because elements of $\mathrm{Aut}(\mathbb{H})$ preserve circles and angles. Hence, since the imaginary axis $T(\mathcal{C})$ is the unique geodesic from $T(P)$ to $T(Q)$, $\mathcal{C}$ must be the unique geodesic from P to Q.
(ii) One way to prove this is to use the Cayley transformation which maps circle in circles and preserves angles. (Do this!) Alternatively one can make an argument similar to the one just seen. To see there is a circle $\mathcal{C}$ as required, consider the transformation $T : z \mapsto \dfrac{z-P}{1-z\bar{P}} \in \mathrm{Aut}(\mathbb{D})$ which carries P to O. Then let $\mathcal{C}'$ be the unique diameter joining O and $T(Q)$ and define $\mathcal{C} = T^{-1}(\mathcal{C}')$. $\square$

Proposition 2.9. *(i) The group* $\mathrm{Isom}^+(\mathbb{H})$ *(resp.* $\mathrm{Isom}^+(\mathbb{D})$*) acts transitively on equidistant pairs of points in* $\mathbb{H}$ *(resp.* $\partial\mathbb{D}$*).*
(ii) The group $\mathrm{Isom}^+(\mathbb{H})$ *(resp.* $\mathrm{Isom}^+(\mathbb{D})$*) acts transitively on ordered triples in* $\partial\mathbb{H}$ *(resp.* $\partial\mathbb{D}$*).*

Proof. We will do the proof in $\mathbb{H}$. To prove the result in $\mathbb{D}$ you can either make direct calculations or apply the Cayley transformation.
(i) We have to show that if $P, P', Q, Q' \in \mathbb{H}$ with $d_{\mathbb{H}}(P, P') = d_{\mathbb{H}}(Q, Q')$, then there exists an isometry $T \in \mathrm{Isom}^+(\mathbb{H})$ such that $T(P) = P'$ and $T(Q) = Q'$.

It will be sufficient to prove this for the case in which $Q = i$ and $Q' = i \cdot e^{d_{\mathbb{H}}(P,P')}$. (Why?) Let S_1 be the map used in the previous proposition which maps the circle $\mathcal{C}$ to the imaginary axis.

Now let S_2 be the map $z \mapsto z/a$ which maps the imaginary axis to itself sending ai to i, so in particular $S_2S_1(P) = i$. Since S_2S_1 is an isometry we have $d_{\mathbb{H}}(S_2S_1(P), S_2S_1(P')) = d_{\mathbb{H}}(P, P')$, so that $d_{\mathbb{H}}(i, S_2S_1(P')) = d_{\mathbb{H}}(P, P')$ and hence $S_2S_1(P') = ie^{\pm d_{\mathbb{H}}(P,P')}$.

If $S_2S_1(P') = i \cdot e^{d_{\mathbb{H}}(P,P')} = Q'$ we are done taking $T = S_2S_1$. Otherwise, apply $S_3 : z \mapsto -1/z$ which fixes $Q = i$ and takes it to i/t for any $t > 0$. (Note that $S_3 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \in \mathrm{SL}(2, \mathbb{R})$.) Then $S_3S_2S_1(P) = Q$ and $S_3S_2S_1(P') = Q'$ so we are done.

(ii) Say $\xi_1, \xi_2, \xi_3 \in \mathbb{R} \cup \infty$. The map

$$T : z \mapsto [z, \xi_1, \xi_2, \xi_3] = \frac{z - \xi_1}{z - \xi_2} \cdot \frac{\xi_3 - \xi_2}{\xi_3 - \xi_1}$$

carries $\xi_1 \mapsto 0, \xi_2 \mapsto \infty, \xi_3 \mapsto 1$. Now $\det T = \dfrac{(\xi_3 - \xi_2)(\xi_1 - \xi_2)}{(\xi_3 - \xi_1)}$, so $\det T > 0$ if and only if the two triples $(0, \infty, 1)$ and (ξ_1, ξ_2, ξ_3) have the same cyclic order on $\hat{\mathbb{R}}$. Then we can normalize to get $T \in \mathrm{SL}(2, \mathbb{R})$. □

Hyperbolic circles

Let $C(z_0, \rho)$ be the hyperbolic circle in $\mathbb{D}$ with centre $z_0 \in \mathbb{D}$ and hyperbolic radius ρ. Since the formula for distance in $\mathbb{D}$ is clearly symmetric under rotation about $0 \in \mathbb{D}$, it follows from the formula (2.1) that $C(0, \rho)$ coincides with the Euclidean circle with centre 0 and Euclidean radius $r = \tanh(\rho/2)$.

Now consider $C(z_0, \rho)$. The isometry $T(z) = \dfrac{z - z_0}{1 - \bar{z}_0 z}$ carries z_0 to 0. Since isometries preserves hyperbolic circles, it follows that $T(C(z_0, \rho)) = C(0, \rho)$ and so $C(z_0, \rho) = T^{-1}(C(0, \rho))$. Since T carries Euclidean circles to Euclidean circles, $C(z_0, \rho)$ is also a Euclidean circle. Notice however that the hyperbolic centre of $C(z_0, \rho)$ does not in general coincide with the Euclidean centre.

Exercise 2.10. *Prove that hyperbolic disks are hyperbolically convex. (To say that a set $X \subset \mathbb{H}$ is hyperbolically convex means that given $x, y \in X$ then the geodesic from x, y is also in X.)*

Armed with these facts about circles, we can now prove that Möbius maps give *all* the orientation preserving isometries of $\mathbb{D}$ and $\mathbb{H}$.

Theorem 2.11.

$$\mathrm{Isom}^+(\mathbb{H}) = \mathrm{Aut}(\mathbb{H}) \qquad \mathrm{Isom}(\mathbb{D}) = \mathrm{Aut}(\mathbb{D}).$$

Proof. From Proposition 2.5 we already know that

$$\mathrm{Isom}^+(\mathbb{H}) \supset \mathrm{Aut}(\mathbb{H}) \qquad \mathrm{Isom}(\mathbb{D}) \supset \mathrm{Aut}(\mathbb{D}).$$

Suppose $T \in \mathrm{Isom}^+ \mathbb{H}$. By composing with a suitable element of $SL(2, \mathbb{R})$ as in the proof of Proposition 2.9 (i), we may assume that T fixes two points P, P' on the imaginary axis $i\mathbb{R}$. Now let $Q \in \mathbb{H}$. Since $d(P, Q) = d(P, T(Q))$, the point $T(Q)$ lies on the circle radius $d(P, Q)$ centre P. Likewise $T(Q)$ lies on the circle radius $d(P', Q)$ centre P'. These two circles are also Euclidean circles, they intersect in precisely two points, one on each side of the imaginary axis. Clearly, one of these two points is Q. Since T preserves orientation, $T(Q)$ must be on the same side of the imaginary axis as Q and so $T(Q) = Q$ (why?) and hence $T = id$. The result in $\mathbb{D}$ is similar, or use the Cayley transform. □

Remark 2.12. *From the above discussion it is easy to deduce:*
(i) $T \in \mathrm{Isom}^+(\mathbb{D})$ is differentiable.
(ii) Every orientation reversing element in $\mathrm{Isom}^-(\mathbb{D})$ is the composition of a reflection $z \mapsto \bar{z}$ with an element in $\mathrm{Isom}^+(\mathbb{D})$.

More formulae for distance

There are several useful ways of writing down the distance between two points without having to integrate along arcs.

1. Cross-ratio form Recall from Chapter 1 that $\mathrm{SL}(2,\mathbb{R})$ preserves cross-ratio. Now we shall see that it can be used to measure distance.

Proposition 2.13.

$$d_{\mathbb{H}}(P,Q) = \log[P',Q,P,Q'] \quad \text{and} \quad d_{\mathbb{D}}(P,Q) = \log[P',Q,P,Q']$$

where P' and Q' are the endpoints on $\partial\mathbb{H}$ or $\partial\mathbb{D}$ of the geodesic that joins P to Q, as shown in Figures 2.5 and 2.6. [Note the importance of the order.]

Proof. Suppose that $T \in \mathrm{Isom}^+(\mathbb{H})$. We know that T is an isometry of $\mathbb{H}$ and preserves cross-ratio. This implies that it maps the geodesic through P, Q to the geodesic through $T(P), T(Q)$, and also that it maps the two endpoints of $[P,Q]$ on $\hat{\mathbb{R}} = \mathbb{R} \cup \infty$ to the endpoints of $[T(P), T(Q)]$ preserving order. Thus it is enough to check for points P and Q on the axis $i\mathbb{R}$. (Why?) If $P = i$ and $Q = ai$ with $a \in \mathbb{R}$, then $P' = 0$ and $Q' = \infty$. In that case

$$d_{\mathbb{H}}(P,Q) = \log a$$

and

$$[P',Q,P,Q'] = \frac{0-ai}{0-i} \cdot \frac{\infty - i}{\infty - ai} = a,$$

as we wanted to prove.

Now use the Cayley transformation to prove the formula in $\mathbb{D}$. □

2. A distance formula in $\mathbb{H}$

A very useful formula for calculating distance in $\mathbb{H}$ is:

Proposition 2.14. *Given two points $z_1, z_2 \in \mathbb{H}$:*

$$\cosh d_{\mathbb{H}}(z_1,z_2) = 1 + \frac{|z_1 - z_2|^2}{2\Im z_1 \Im z_2}.$$

Proof. To prove this formula it is enough to check that both sides are invariant under isometries and then to check that the formula works when $z_1, z_2 \in \mathbb{R}$. All of these are easy to do. □

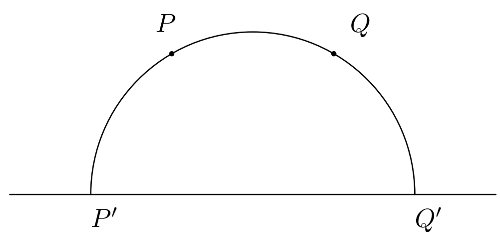

Figure 2.5: Cross ratio form of the distance in $\mathbb{H}$

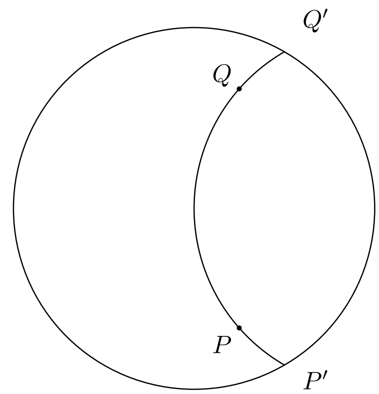

Figure 2.6: Cross ratio form of the distance in $\mathbb{D}$

This formula can be arranged in other ways which are sometimes useful:

$$d_{\mathbb{H}}(z_1, z_2) = \ln \frac{|z_1 - \bar{z}_2| + |z_1 - z_2|}{|z_1 - \bar{z}_2| + |z_1 - z_2|} \quad ; \quad \sinh(d_{\mathbb{H}}(z_1, z_2)/2) = \frac{|z_1 - z_2|}{2(\Im z_1 \Im z_2)^{\frac{1}{2}}};$$

$$\cosh(d_{\mathbb{H}}(z_1, z_2)/2) = \frac{|z_1 - \bar{z}_2|}{2(\Im z_1 \Im z_2)^{\frac{1}{2}}} \quad ; \quad \tanh(d_{\mathbb{H}}(z_1, z_2)/2) = \frac{|z_1 - z_2|}{|z_1 - \bar{z}_2|}.$$

It is an easy exercise to prove that the four equations are equivalent to each other and to the equations of Proposition 2.14. There are also analogous formulae in $\mathbb{D}$:

Corollary 2.15. *Given two points $z_1, z_2 \in \mathbb{D}$ we have:*

$$d_{\mathbb{D}}(z_1, z_2) = \ln \frac{|1 - z_1\bar{z}_2| + |z_1 - z_2|}{|1 - z_1\bar{z}_2| - |z_1 - z_2|} \quad ; \quad \cosh^2(d_{\mathbb{D}}(z_1, z_2)/2) = \frac{|1 - z_1\bar{z}_2|^2}{(1 - |z_1|^2)(1 - |z_2|^2)};$$

$$\sinh^2(d_{\mathbb{D}}(z_1, z_2)/2) = \frac{|z_1 - z_2|^2}{(1 - |z_1|^2)(1 - |z_2|^2)} \quad ; \quad \tanh(d_{\mathbb{D}}(z_1, z_2)/2) = \frac{|z_1 - z_2|}{|1 - z_1\bar{z}_2|}.$$

Proof. This is an exercise using the previous results and the fact that $d_{\mathbb{D}}(z_1, z_2) = d_{\mathbb{H}}(C^{-1}(z_1), C^{-1}(z_2))$ where $C : \mathbb{H} \longrightarrow \mathbb{D}$ is the Cayley transformation, which follows from Proposition 2.6. For details, see [10] p. 131. ∎

2.2 Similarities and differences between hyperbolic and Euclidean geometry

2.2.1 Circles

As we saw above, hyperbolic circles are off-centre Euclidean circles. In particular, the hyperbolic circle centre $O \in \mathbb{D}$ and radius ρ coincides with the Euclidean circle centre $O \in \mathbb{D}$ and Euclidean radius $r = \tanh \rho/2$.

Lemma 2.16. *The circumference of a hyperbolic circle of radius ρ is $2\pi \sinh(\rho)$. The area of a hyperbolic disc of radius ρ is $4\pi \sinh^2(\frac{\rho}{2})$.*

Proof. The hyperbolic length of the circumference of a circle with radius ρ is

$$\int_0^{2\pi} \frac{2r}{1-r^2} d\theta = \frac{4\pi r}{1-r^2}$$

where $r = \tanh(\frac{\rho}{2})$. But $1 - r^2 = \cosh^{-2}(\frac{\rho}{2})$, so

$$\text{Circumference} = 4\pi \sinh(\frac{\rho}{2}) \cosh(\frac{\rho}{2}) = 2\pi \sinh(\rho).$$

The hyperbolic area of the disc with radius ρ is

$$\int_{t=0}^{r} \int_0^{2\pi} \frac{4t}{(1-t^2)^2} dt d\theta = 8\pi \left[\frac{1}{2(1-t^2)}\right]_0^r = 4\pi \left(\frac{1}{1-r^2} - 1\right) = \frac{4\pi r^2}{1-r^2}.$$

So

$$\text{Area} = 4\pi \sinh^2(\frac{\rho}{2}).$$

□

Remark 2.17. *Observe the very important fact that both the circumference and the area grow exponentially with the radius.*

Exercise 2.18 (Circles in $\mathbb{H}$)**.** *We know that a circle in $\mathbb{H}$ must be an off-centre Euclidean circle. If the circle has radius ρ and centre i, you can see easily that it meets imaginary axis at points ie^{ρ} and $ie^{-\rho}$, as shown in Figure 2.7 . Clearly it must be symmetric about the imaginary axis. What are its Euclidean centre and radius?*

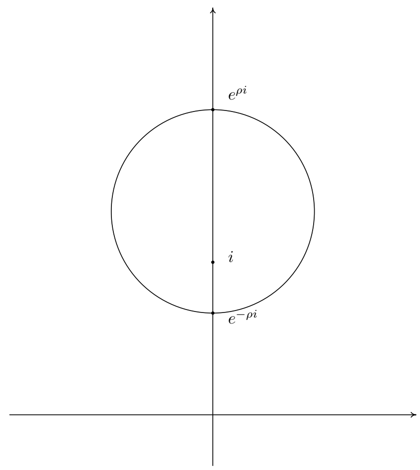

Figure 2.7: Hyperbolic circle centre i in $\mathbb{H}$

2.2.2 Angle sum of a triangle

One of the most important differences between Euclidean and hyperbolic geometry is the angle sum of a triangle:

Lemma 2.19. *The angle sum of a hyperbolic triangle is strictly less than π.*

Proof. Put the hyperbolic triangle Δ in $\mathbb{D}$ with one vertex at the centre O. The sides through O are radii. If P, P' are the other two vertices, it is easy to see that the Euclidean triangle with vertices OPP' strictly contains Δ, and in particular, that the $\angle OPP', \angle OP'P$ in Δ are less than the corresponding Euclidean angles, from which the result follows. □

Corollary 2.20. *The angle sum of a hyperbolic polygon with at most n sides is strictly less than $(n-2)\pi$.*

Notice that this includes the case of triangles (or polygons) with one or more vertices on the boundary at infinity. At any such vertex, the angle is zero, because if two lines meet at a point $P \in \partial\mathbb{H}$, then they are both orthogonal to $\partial\mathbb{H}$ and hence the angle between them is zero.

We shall find the precise relation between area and angle sum later.

2.2.3 Perpendicular projection and the common perpendicular.

In Euclidean geometry, there is a unique perpendicular projection from a point to a line. The same is true in hyperbolic space:

Theorem 2.21. *There exists a unique perpendicular from a point P to a line $\mathcal{L}$. This perpendicular minimises the distance from P to $\mathcal{L}$.*

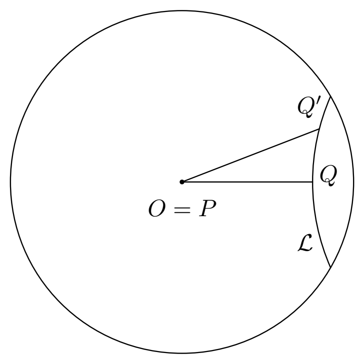

Figure 2.8: The hyperbolic perpendicular from the point P to the line $\mathcal{L}$.

Proof. Place P at $O \in \mathbb{D}$. Referring to Figure 2.8, it is clear from Euclidean geometry that there is a unique radial line PQ perpendicular to the hyperbolic line $\mathcal{L}$. The relation $d_{\mathbb{D}}(P,Q) < d_{\mathbb{D}}(P,Q')$ follows since from Euclidean geometry, any other point $Q' \in \mathcal{L}$ lies outside the hyperbolic circle centre 0 and radius $|PQ|$. □

Suppose that $\mathcal{L}$ and $\mathcal{L}'$ are are parallel lines in Euclidean space. Then they have infinitely many common perpendiculars, each of which minimises the distance between them. By contrast, in hyperbolic space:

Theorem 2.22. *Let $\mathcal{L}$ and $\mathcal{L}'$ be disjoint lines which do not meet at ∞. Then $\mathcal{L}, \mathcal{L}'$ have a **unique** common perpendicular which minimizes the distance between them.*

Proof. • *Existence*: This time we arrange line $\mathcal{L}$ to be the imaginary axis in $\mathbb{H}$. We may suppose that $\mathcal{L}'$ is a Euclidean semicircle as shown in Figure 2.9.

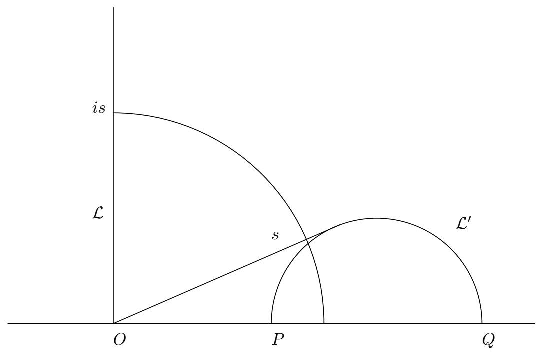

Figure 2.9: The circle centre O and radius s is the common perpendicular from $\mathcal{L}$ to $\mathcal{L}'$.

(Why?) It follows from Euclidean geometry that the circle with centre O and radius s, where $s^2 = |OP| \cdot |OQ|$, cuts $\mathcal{L}'$ orthogonally. • *Uniqueness*: Any other common perpendicular would produce either a quadrilateral whose angle sum is 2π or a triangle of angle sum larger then π, as shown in the figure. Both of these cases are impossible by Lemma 2.19 and its corollary.
• *The common perpendicular minimizes distance.* Denote the common perpendicular by $\mathcal{M}$ and let X, X' be the points where $\mathcal{M}$ meets $\mathcal{L}, \mathcal{L}'$ respectively. Let Y, Y' be any other two points on $\mathcal{L}, \mathcal{L}'$ respectively. The result follows immediately from the next theorem, on noting that the perpendicular projections from Y, Y' onto $\mathcal{M}$ are exactly the points X, X'. Theorem 2.23 says that $d_{\mathbb{H}}(Y, Y') > d_{\mathbb{H}}(X, X')$.

□

Theorem 2.23. *Perpendicular projection onto a line $\mathcal{M}$ strictly decrease distances. More precisely, suppose that $P, Q \notin \mathcal{M}$ and let P', Q' denote the perpendicular projections of points P, Q onto $\mathcal{M}$. Then $d(P, Q) > d(P', Q')$.*

Proof. In $\mathbb{H}$, move $\mathcal{M}$ to the imaginary axis by sending the points P', Q' to i and $e^R i$ respectively, so that $d_{\mathbb{H}}(P', Q') = R$. The line PP' meets $\mathcal{M}$ orthogonally and hence P lies on the geodesic through i perpendicular to the imaginary axis, so $P = e^{i\theta}$ for some $\theta \in (0, \pi)$. Similarly we can write $Q = e^R e^{i\phi}$ for some $\phi \in (0, \pi)$.

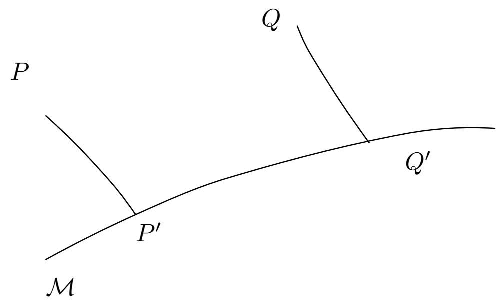

Figure 2.10: Perpendicular projection decreases distances: $|P'Q'| < |PQ|$.

The distance formula in Proposition 2.14 gives

$$\begin{aligned}
\cosh d_{\mathbb{H}}(P,Q) &= 1 + \frac{|e^{i\theta} - e^R e^{i\phi}|^2}{2\sin\theta e^R \sin\phi} \\
&= 1 + \frac{1 + e^{2R} - 2e^R(\cos(\theta-\phi)}{2e^R \sin\theta \sin\phi} \\
&\geq 1 + \frac{(1-e^R)^2}{2e^R\sin\theta\sin\phi} \geq 1 + 2\sinh^2(R/2).
\end{aligned}$$

Using $\cosh d = 1 + 2\sinh^2 d/2$ this gives

$$\sinh^2 d_{\mathbb{H}}(P,Q) \geq \sinh^2(R/2) = \sinh^2 d_{\mathbb{H}}(P',Q')$$

so that $d_{\mathbb{H}}(P,Q) \geq d_{\mathbb{H}}(P',Q')$ with equality if and only if $P = P'$ and $Q = Q'$.

We remark that using the angle of parallelism formula, see Lemma 2.27 below, we have $\cosh d_{\mathbb{H}}(P,P') = \dfrac{1}{\sin\theta}$ and $\cosh d_{\mathbb{H}}(Q,Q') = \dfrac{1}{\sin\phi}$. Hence our computations actually show

$$\sinh^2 d_{\mathbb{H}}(P,Q) \geq \sinh^2(R/2)\cosh d_{\mathbb{H}}(P,P')\cosh d_{\mathbb{H}}(Q,Q').$$

□

2.2.4 Lines meeting on $\partial\mathbb{H}$

Theorem 2.22 shows that if two lines $\mathcal{L}$ and $\mathcal{L}'$ do not meet even at infinity, then there is a non-zero shortest distance between them which is realised by the unique common perpendicular. If the lines have the same endpoint on $\partial\mathbb{H}$ or $\partial\mathbb{D}$, however, then it is not hard to show that the infimum of the distance between points on the two lines vanishes, in symbols:

$$\inf\{d(P, P') : P \in \mathcal{L}, P' \in \mathcal{L}'\} = 0.$$

The calculation is done by arranging the common point of $\mathcal{L}$ and $\mathcal{L}'$ to be $\infty \in \mathbb{H}$, see Figure 2.11. If we take the two points P, P' to have the same imaginary part h, then

$$d_{\mathbb{H}}(P, P') \leq \int_0^1 \frac{dx}{h} = \frac{1}{h} \longrightarrow 0 \text{ as } h \longrightarrow \infty.$$

Exercise 2.24. *In fact* $d_{\mathbb{H}}(P, P') < 1/h$. *Why? Can you calculate the actual value?*

We say the lines $\mathcal{L}$ and $\mathcal{L}'$ *meet at infinity*. Because both meet $\partial\mathbb{H}$ orthogonally, the angle between such lines is zero. Notice that in this case, there is no geodesic which actually realises the infimum.

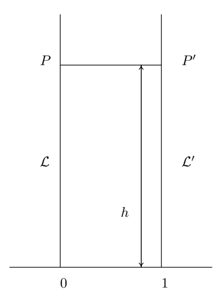

Figure 2.11: The lines $\mathcal{L}$ and $\mathcal{L}'$ meet at ∞; $d_{\mathbb{H}}(P, P') = 1/h$.

If a 'triangle' or 'polygon' has two sides which meet at ∞ we call the meeting point an *ideal vertex*. The vertex is not in $\mathbb{H}$ but we can still identify it as a definite point on $\partial\mathbb{H}$ which can be transported around by $SL(2, \mathbb{R})$. If all vertices are ideal, the triangle (resp. polygon) is called an *ideal triangle* (resp. polygon).

Exercise 2.25. *Show that if* Δ, Δ' *are two ideal triangles in* $\mathbb{H}$*, there is an isometry carrying on onto the other. Is this isometry unique?*

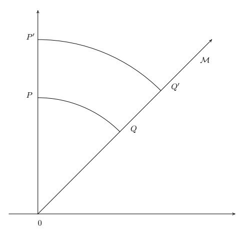

Figure 2.12: The set of points equidistant from the imaginary axis.

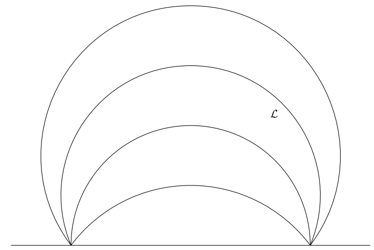

Figure 2.13: Points equidistant from the line $\mathcal{L}$. These lines are sometimes called **hypercycles**.

Let us summarize these results, contrasting with Euclidean geometry. In Euclidean geometry, either two lines meet and they have no common perpendicular, or they are parallel and they have infinitely many common perpendiculars. In hyperbolic geometry, by contrast:
(a) if two lines meet, then as in Euclidean geometry they have no common perpendicular, for this would give a triangle with angle sum $> \pi$.
(b) if two lines are "parallel", that is they meet at ∞, there is still no common perpendicular because that would give a triangle with angle sum equal to π, contradicting Lemma 2.19. The lines become asymptotically close as they approach ∞ and the angle between them is zero.
(c) if the two lines are "ultra-parallel", that is they don't meet and they aren't parallel, then there exists a unique common perpendicular which minimizes the distance between them.

2.2.5 Points equidistant from a given line

In Euclidean geometry, the line parallel to $\mathcal{L}$ through the point $P \notin \mathcal{L}$ consists of all points Q such that $d(Q, \mathcal{L}) = d(P, \mathcal{L})$. What about the set of points equidistant from a hyperbolic line?

Theorem 2.26. *Let $\mathcal{L}$ be a hyperbolic line. The set of all points at a fixed distance from $\mathcal{L}$ is a Euclidean circle with the same endpoints as $\mathcal{L}$ on $\partial\mathbb{H}$ (or $\partial\mathbb{D}$), making an angle θ with $\partial\mathbb{H}$ (or $\partial\mathbb{D}$).*

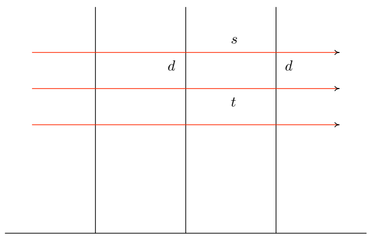

Figure 2.14: Horocycles based at ∞. Note the distance relationship $s = te^{-d}$.

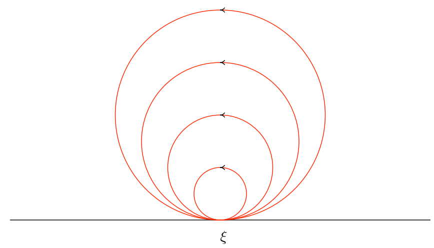

Figure 2.15: Horocycles based at the point $\xi \in \mathbb{R}$.

Proof. First consider the case when $\mathcal{L}$ is the imaginary axis in $\mathbb{H}$. Let $\mathcal{M}$ be a **Euclidean** line through 0, making an angle $\theta \in (0, \pi/2)$ with $\partial\mathbb{H}$. The 1-parameter family of maps $\{z \mapsto \lambda z : \lambda \in \mathbb{R}^+\}$ fixes $\mathcal{M}$ and $\mathcal{L}$ and acts transitively on each by isometries. In particular, referring to Figure 2.12, $d_{\mathbb{H}}(P, Q) = d_{\mathbb{H}}(P', Q')$.

In the general case, transporting the points $0, \infty$ to the points $\xi, \eta \in \partial\mathbb{H}$, we see that the locus of points equidistant from a given line $\mathcal{L}$ consists of points on a circle with the same endpoints ξ, η as $\mathcal{L}$ on $\partial\mathbb{H}$ and which meets $\partial\mathbb{H}$ at angle θ. □

2.2.6 Horocycles

A horocycle is the limit of a circle as the centre approaches $\partial\mathbb{H}$. In our models, they are Euclidean circles tangent to the boundary at infinity. Thus in $\mathbb{H}$ they are either horizontal lines if the tangency point $\xi = \infty$, or circles tangent to the real axis if $\xi \in \mathbb{R}$, see Figures 2.15 and 2.14. As we shall see later, horocycles will play an important role.

Suppose that points P, P' lie on the same horocycle. It is useful to compare the distance ℓ between P, P' measured along the horocycle and the geodesic distance. To do this, as in Figure 2.11 we can arrange the horocycle to be a horizontal line at Euclidean height $h > 0$ above the real axis in $\mathbb{H}$ and by translating and scaling assume that $P = ih, P' = ih + 1$. (Why?)

Integrating along the horocycle path from P to P' as above we have

$$\ell = \int_0^1 \frac{dx}{h} = 1/h.$$

On the other hand, the distance formula for $d = d_{\mathbb{H}}(P, P')$ gives

$$\cosh d = 1 + \frac{|ih - (ih+1)|^2}{2h^2} = 1 + \frac{1}{2h^2}.$$

Using $\cosh d = 1 + 2\sinh^2 d/2$ we find $\sinh d/2 = \dfrac{1}{2h}$ so that $\ell = 2\sinh d/2 < e^{d/2}$. This shows that it is exponentially further to travel along the horocycle than to go along the geodesic path (which penetrates inside the horoball).

We also remark that a horocycle can be thought of as the limit of a circles of fixed radius R whose centres tend to ∞. To see this, recall that the hyperbolic circle centre $ai \in \mathbb{H}$ is the Euclidean circle whose diameter is the interval between the points $ae^R i$ and $ae^{-R} i$. As $a \longrightarrow 0$ with R fixed, the centre tends to $\partial\mathbb{H}$ and the circle tends to a horocycle tangent to $\partial\mathbb{H}$ at 0.

2.2.7 Angle of parallelism

If a point P is at (hyperbolic) distance d from a line $\mathcal{L}$, there is a limiting value θ to the angle made by lines $\mathcal{L}'$ through P not meeting $\mathcal{L}$? The angle θ is classically called the "angle of parallelism" . The answer is given by the following lemma, which describes the trigonometric relations in a triangle with angles $0, \frac{\pi}{2}, \theta$.

Lemma 2.27 (Angle of parallelism)**.** *In a triangle with angles* $0, \frac{\pi}{2}, \theta$ *we have:*

$$\cosh d = \csc\theta; \qquad \sinh d = \cot\theta; \qquad \tanh d = \cos\theta.$$

Proof. Note that there only two parameters: θ and d, as shown in Figure 2.16. Put the triangle in $\mathbb{H}$ so that $\mathcal{L}$ is the imaginary axis with the vertex Q at i, see Figure 2.17. By the formula in Proposition 2.14, we have

$$\cosh d = \cosh d_{\mathbb{H}}(i, P) = \cosh d_{\mathbb{H}}(i, e^{i\theta}) = 1 + \frac{|i - e^{i\theta}|^2}{2\Im i \Im e^{i\theta}} = \frac{1}{\sin\theta} = \csc\theta,$$

so $\cosh d = \csc\theta$. The other two formulae follow immediately by the relations between hyperbolic trigonometric functions. In fact,

$$\sinh d = \sqrt{\sin^{-2}(\theta) - 1} = \cot\theta$$

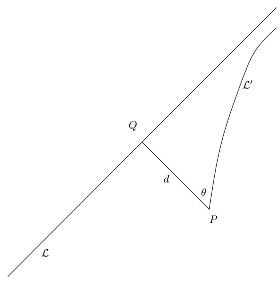

Figure 2.16: The angle of parallelism. Note that there only two parameters: θ and d.

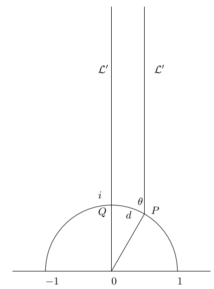

Figure 2.17: Standard position for the angle of parallelism

and

$$\tanh d = \frac{\sinh d}{\cosh d} = \cos\theta.$$

Observe that this argument also shows also the existence of a perpendicular segment from a point to a line. □

2.2.8 Schweikart's constant

There is an upper bound to the length of the altitude of any hyperbolic isosceles right-angled triangle. This is called **Schweikart's constant**. The fact there is such a bound is one of the important differences between hyperbolic and Euclidean geometry.

Let try to calculate the constant. As shown in Figure 2.18, put the triangle with the right angled vertex at 0. The figure shows the case in which the other two vertices are on $\partial\mathbb{D}$. Any other right angled isoceles triangle is contained inside the one shown, so this is the limiting case. (Why?) It is an easy Euclidean calculation to find that $|OX| = \sqrt{2} - 1$. (Check it as an exercise!). So we obtain **Schweikart's constant**:

$$d_{\mathbb{D}}(O, X) = \log(1 + \sqrt{2}).$$

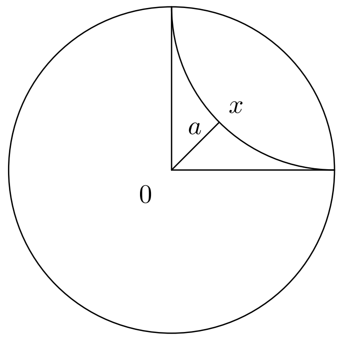

Figure 2.18: Computing Schweikart's constant.

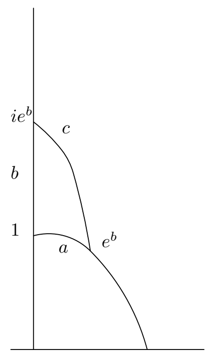

Figure 2.19: Standard position for a right angled triangle in $\mathbb{H}$.

2.2.9 Pythagoras' theorem

Suppose we have a right angled triangle with side lengths a, b, c where c is the hypotenuese. What is the relation between a, b and c? Move the triangle to the standard position shown in Figure 2.19. Using the calculations done in Section 2.3.2, we have the relation $\cosh a = \dfrac{1}{\sin\theta}$ and, using the formula proved in the Subsection 2.2.3, it is an easy calculation to show that

$$\cosh c = 1 + \frac{|ie^b - e^{i\theta}|^2}{2\Im ie^b \Im e^{i\theta}} = \frac{1}{\sin\theta}\cdot\left(\frac{e^b + e^{-b}}{2}\right) = \frac{1}{\sin\theta}\cosh b = \cosh a \cosh b.$$

So we obtain the hyperbolic version of **Pythagoras' theorem**:

$$\cosh c = \cosh a \cosh b.$$

This leads to another crucial difference with Euclidean geometry. In *hyperbolic* geometry the above formula gives:

$$\cosh(a+b) = \cosh a \cosh b + \sinh a \sinh b \leqslant 2 \cosh a \cosh b = 2 \cosh c =$$

$$\cosh(\text{arccosh} 2) \cosh c \leqslant \cosh(c + \text{arccosh} 2)$$

and so

$$a + b < c + \text{arccosh} 2,$$

independently of a, b and c.

By contrast, in *Euclidean* geometry, the relation between the sides of a right-angled triangle is:

$$a^2 + b^2 = c^2 \Rightarrow (a+b)^2 = c^2 + 2ab$$

giving

$$\Rightarrow a + b \gg c \text{ as } a, b \longrightarrow \infty.$$

This says that while in Euclidean space, it is much quicker to cut across the diagonal of a field, in hyperbolic geometry it makes almost no difference if you stick to the edge.

We can see another version of the same phenomenon referring back to Theorem 2.23 about perpendicular projection onto a line. Suppose (see Figure 2.10) that points P, Q are equal distances d from a line $\mathcal{M}$, let P', Q' be their perpendicular projections onto $\mathcal{M}$. We proved in Theorem 2.23 that $\sinh d(P,Q)/2 = \sinh d(P',Q')/2 \cosh d$. For reasonably large values of $d(P',Q')$ this gives $d(P,Q) \sim d(P',Q') + 2d$, which can be interpreted as saying that the geodesic path from P to Q roughly tracks the path $PP' \cup P'Q' \cup Q'Q$.

2.2.10 Area of a triangle

Theorem 2.28. *The area of a triangle with angles α, β, γ is $\pi - (\alpha + \beta + \gamma)$.*

Remark 2.29. The area formula agrees with the Gauss-Bonnet Formula which says that for any surface M with Riemannian metric,

$$\int_M K dA + \int_{\partial M} k_g ds = 2\pi\chi(M)$$

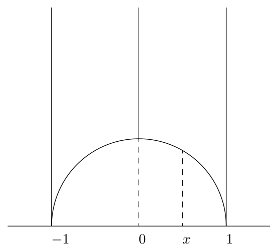

Figure 2.20: Standard position to compute the area of an ideal triangle.

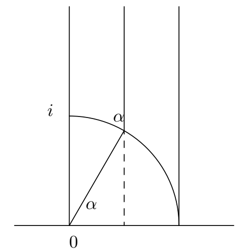

Figure 2.21: Area of a triangle with angles $\alpha, \pi/2, 0$.

where K is the curvature, k_g is the geodesic curvature along the boundary ∂M and $\chi(M)$ is the Euler characteristic. In our case we know $K \equiv -1$. Since the boundary is piecewise geodesic, the curvature is concentrated at the vertices and the term $\int_{\partial M} k_g ds$ reduces to the sum of the *exterior* angles as we travel round ∂M. Finally, M is topologically a disk, so $\chi(M) = 1$. This gives

$$\int (-1) dA + \sum (\pi - \alpha_i) = 2\pi$$

which reduces to the area formula above.

Proof. Method 1:
• We first compute area of an "ideal triangle" with all vertices at ∞ (hence all angles are 0). Since $\mathrm{Aut}(\mathbb{H})$ acts transitively on ordered triples of points on $\hat{\mathbb{R}}$, we may assume that the vertices are -1, 1 and ∞, see Figure 2.20. Then the area is:

$$\text{Area} = 2\int_0^1 \left(\int_{\sqrt{1-x^2}}^{\infty} \frac{dy}{y^2} \right) dx = 2\int_0^1 \frac{dx}{\sqrt{1-x^2}} = 2\left[\arcsin x\right]_0^1 = 2 \cdot \frac{\pi}{2} = \pi.$$

• We can equally compute area of a triangle with angles $\pi/2, \alpha$ and 0. As can be checked from Figure 2.21, this triangle can be placed with vertices at $\infty, i, e^{i\alpha}$.

$$\text{Area} = \int_0^{\cos\alpha} \left(\int_{\sqrt{1-x^2}}^{\infty} \frac{dy}{y^2} \right) dx = \left[\arcsin x\right]_0^{\cos\alpha} = \frac{\pi}{2} - \alpha.$$

• To compute area of a triangle with angles α, β and 0, we juxtapose two triangles of angles $\pi/2, \alpha$ and 0 and $\pi/2, \beta$ and 0 as shown in Figure 2.22.

(If $\alpha > \pi/2$, then we can find the area by noting that a triangles of angle $\alpha, 0, \pi/2$ and $\pi - \alpha, 0, \pi/2$ can be juxtaposed to give a triangle of angles $0, \pi/2, 0$.)

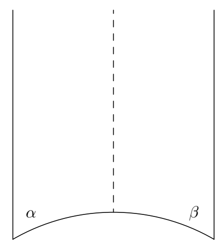

Figure 2.22: Area of a triangle with angles $\alpha, \beta, 0$.

$$\text{Area} = \left(\frac{\pi}{2} - \alpha\right) + \left(\frac{\pi}{2} - \beta\right) = \pi - (\alpha + \beta).$$

• Finally we find the area of a triangle with angles α, β and γ at vertices A, B, C as follows. Extend the side through AC to meet $\partial\mathbb{H}$ at X. Then

$$\text{Area } ABC = \text{Area } ABX - \text{Area } BCX.$$

Let $\delta = \angle CBX$. Then

$$\text{Area } ABX - \text{Area } BCX = (\pi - (\alpha + \beta + \delta)) - (\pi - (\delta + (\pi - \gamma))) = \pi - (\alpha + \beta + \gamma)$$

and the result follows.
Method 2:
• It follows from the fact that $\text{Isom}(\mathbb{H})$ acts transitively on ordered triples of points on $\partial\mathbb{H}$, that all the ideal triangles (ie triangles with all vertices on $\partial\mathbb{H}$) are congruent. The first part of Method 1 gives that the area of such a triangle is π.
• Let the area of a triangle with angles $\theta, 0$ and 0 be $\pi - f(\theta)$. We have that

$$f(0) = 0, \qquad f(\pi) = \pi$$

and

$$\pi - f(\theta) + \pi - f(\phi) = \pi - f(\theta + \phi) + \pi.$$

Hence

$$f(\theta) + f(\phi) = f(\theta + \phi),$$

ie f is linear.
• Assuming the continuity of f, this gives

$$f(\theta) = k \cdot \theta$$

for some $k \in \mathbb{R}$. Since $f(\pi) = \pi$, we obtain $k = 1$. Hence the area of a triangle with angles $\theta, 0$ and 0 is $(\pi - \theta)$.
• Given a general triangle with vertices ABC, extend the oriented sides AB, BC, CA to meet $\partial\mathbb{H}$ in points X, Y, Z say. Then

$$\text{Area } ABC = \text{Area } XYZ - \text{Area } AXZ - \text{Area } BXY - \text{Area } CYZ.$$

Check that the result follows.

□

2.2.11 Congruent triangles

Definition 2.30. *Two triangles ABC and $A'B'C'$ are* ***congruent*** *if there exists a $T \in \text{Isom}(\mathbb{H})$ such that $T(A) = A', T(B) = B'$ and $T(C) = C'$. (Notice that in this definition, T may or may not preserve orientation.)*

Recall the conditions for Euclidean congruence between two triangles:
(a) SAS (side-angle-side) means the two triangles have two sides of the same length together with the included angle;
(b) SSS; (three sides of equal length)
(c) SAA; (one side and two angles equal)
(d) RHS (right angle-hypotenuese-side).

Let us try these conditions in the hyperbolic case. Let's work in $\mathbb{D}$. Without loss of generality, we can put one vertex at O and make one side a horizontal diameter. (Why?) Then:
(a) SAS works as in Euclidean geometry;
(b) SSS: the hyperbolic circles with centres A and B and radius $|AC|$ and $|BC|$ where $|AC| > |BC|$ are also Euclidean circles, so they intersect in exactly two points. Thus the triangle is determined uniquely up to an orientation preserving isometry, as in the Euclidean case.
(c) SAA works as in Euclidean geometry. (d) RHS: works as in Euclidean geometry. (Check as an exercise using the fact that there exists a unique perpendicular from a point to a line).

In hyperbolic geometry we have also a fourth condition for triangles to be congruent: AAA. In fact we have:

Theorem 2.31. *(i) Any two triangles with the same internal angles are congruent.*
(ii) For any α, β, γ with $\alpha + \beta + \gamma < \pi$ and $\alpha, \beta, \gamma \geqslant 0$, there exists a triangle with these angles. (Recall that an angle is 0 $\iff$ the vertex is an "ideal vertex" at ∞)

Proof. (i) Consider two triangles of angles α, β, γ at vertices A, B, C and A', B', C'. Applying an isometry, we may arrange the triangles so that $A = A'$ and so that the sides AB, AB' are contained in a common line through A, as are AC, AC'.

Suppose without loss of generality that $|AB| \leq |AB'|$. If $|AC| \leq |AC'|$ then either triangle ABC is strictly contained inside triangle $AB'C'$, which is impossible because they have the same area, or $B = B', C = C'$.

If $|AC| > |AC'|$ then the lines BC and $B'C'$ cross at a point X say. Then the angle sum of triangle BXB' is at least $\beta + (\pi - \beta)$, which is impossible.

(ii) We already constructed an $\alpha, \beta, 0$ triangle in the proof of Theorem 2.28 above, see also Figure 2.21. We can construct an α, β, γ triangle using a *continuity argument* as follows. Place the vertex A at $O \in \mathbb{D}$ and let B be the other finite vertex. If B' is any point between A and B, the line through B' making an angle β with AB (on the same side as BC) cuts the third side AC in a point C' say. Let θ be the internal angle at C'.

As B' moves towards A, the area of triangle $AB'C'$ decreases. Hence since the angles at A, B' are fixed by construction, θ must be strictly increasing. When $B' = B$ then $C' = C$ and $\theta = 0$; when $B' \longrightarrow A$ we have $C' \longrightarrow A$ also. (Why?) In this case the triangle $AB'C'$ becomes almost Euclidean so that $\theta \longrightarrow \pi - (\alpha + \beta)$. Thus by continuity, there must be a position of B' for which θ takes any value in $[0, \pi - (\alpha + \beta))$. □

Remark 2.32. *Another way to see this is to use the second cosine rule:*

$$\cosh c = \frac{\cos\gamma + \cos\alpha\cos\beta}{\sin\alpha\sin\beta},$$

see Examples sheet 4. This gives the length of one side given the three angles, and we can now construct the triangle using ASA.

2.2.12 Polygons

Recall that for an N-sided Euclidean polygon, the sum of the internal angles α_i is $\sum_{i=1}^{N} \alpha_i = (N-2)\pi$.

Theorem 2.33. *The area of an N-sided polygon is $(N-2)\pi - \sum_{i=1}^{N} \alpha_i$.*

Proof. Divide the polygon into $N-2$ triangles by joining one vertex to all of the others. The sum of their internal angles is exactly $\sum_{i=1}^{N} \alpha_i$. □

Example 2.34. *Observe that the angle of a regular octagon in the Euclidean geometry is $3\pi/4$. (Why?) By contrast, there exists a regular octagon all of whose interior angles are $\pi/4$.*

• *Method 1*: Construct a triangle with angles $\pi/4, \pi/8, \pi/8$ (How?) and put 8 of these together round the central vertex $\pi/4$.
• *Method 2*: Use a continuity argument. In the disk $\mathbb{D}$ consider the ideal octagon whose vertices are at 8 equally spaced points round $\mathbb{D}$. Move the vertices inward symmetrically along the radii. It is not hard to see that the vertex angles increase from the initial value 0 to the limiting value of a Euclidean polygon, namely $3\pi/4$. This value is never reached but is approached as the distance from O to the vertices tends to zero. Since the angle is clearly continuous, there must be a point at which it takes on any value in the interval $(0, 3\pi/4)$, in particular, at which it has the value $\pi/4$.

2.2.13 Hyperbolic trigonometry

There are formulae in hyperbolic geometry analogous to those in Euclidean geometry. For elegant proofs using the Minkowski model, se [9] or [14]. here we sketch the proofs in the disc and upper half plane models; for more details see [10].

Right angled triangles The trigonometric formulae for a right angled triangle ABC, with sides of hyperbolic lengths a, b, c opposite angles $\alpha, \beta, \frac{\pi}{2}$ at vertices A, B, C respectively are:

$$\cos\beta = \tanh a/\tanh c, \quad \sin\beta = \sinh b/\sinh c, \quad \tan\beta = \tanh b/\sinh a.$$

Here are steps to verify this:

1. The hyperbolic line which meets the horizontal radius at a point hyperbolic distance $a > 0$ from $O \in \mathbb{D}$ orthogonally is an arc of circle orthogonal to $\partial\mathbb{D}$. Prove that the (Euclidean) radius of this circle is $(\sinh a)^{-1}$.

2. Put the vertex B with angle β at $O \in \mathbb{D}$ so that vertex C is at distance a from O along a horizontal diameter. Let X be the centre of the (Euclidean) circle through A orthogonal to $\partial\mathbb{D}$. Use the Euclidean cosine formula in triangle OAX to prove that $\cos\beta = \tanh a / \tanh c$.

3. Use the hyperbolic Pythagoras theorem to prove the other two identities.

The hyperbolic sine and cosine laws Notice that there are two cosine rules, one analogous to the Euclidean cosine rule, and one which gives the length of a side of a triangle in terms of its angles, which is of course impossible in the Euclidean case.

Consider a general hyperbolic triangle with angles α, β, γ and opposite sides of lengths a, b, c. Use the above formulae in for a right angled triangle to prove:

The hyperbolic sine law

$$\frac{\sinh a}{\sin\alpha} = \frac{\sinh b}{\sin\beta} = \frac{\sinh c}{\sin\gamma}.$$

Hint: drop the perpendicular from one vertex to the opposite side.

The first hyperbolic cosine law

$$\cos\gamma = \frac{\cosh a \cosh b - \cosh c}{\sinh a \sinh b}.$$

The second hyperbolic cosine law

$$\cos c = \frac{\cos\gamma + \cos\alpha\cos\beta}{\sin\alpha\sin\beta}.$$

Hint: Let the perpendicular from C to AB meet AB in D. Let CD have length h and let the angle at C in triangle CAD be γ_1. First verify the identity $\sin\gamma_1 \cosh h = \cos\alpha$.

Exercise 2.35. *Research the corresponding formulae for spherical trigonometry.*

Chapter 3

Isometries of $\mathbb{D}$ and $\mathbb{H}$

In this chapter we are going to study the behaviour and dynamical properties of isometries of $\mathbb{D}$ and $\mathbb{H}$. It is natural to study such transformations up to conjugation in the isometry groups, because, as we shall see shortly, conjugate elements behave in a “similar way”. We shall begin with general results about $\operatorname{Aut}\hat{\mathbb{C}}$ and then specialise to $\operatorname{Aut}(\mathbb{D})$ and $\operatorname{Aut}(\mathbb{H})$.

3.1 Classification of elements of $\operatorname{Aut}(\hat{\mathbb{C}})$.

Definition 3.1. *Two elements T and $\hat{T}$ in $\operatorname{Aut}(\hat{\mathbb{C}})$ are* ***conjugate*** *if there exists an element $S \in \operatorname{Aut}(\hat{\mathbb{C}})$ such that $\hat{T} = STS^{-1}$.*

Thus we have the following commutative diagram:

$$\begin{array}{ccc} \hat{\mathbb{C}} & \xrightarrow{T} & \hat{\mathbb{C}} \\ S \downarrow & & \downarrow S \\ \hat{\mathbb{C}} & \xrightarrow{\hat{T}} & \hat{\mathbb{C}} \end{array}$$

What this means is that after applying S to $\hat{\mathbb{C}}$, the map $\hat{T}$ behaves like T. For example:

Example 3.2. *If z_0 is a fixed point of T, that is $T(z_0) = z_0$, then $S(z_0)$ is a fixed point of $\hat{T}$.*

CHECK:

$$\hat{T}\left(S(z_0)\right) = STS^{-1}\left(S(z_0)\right) = ST(z_0) = S(z_0). \ \square$$

Definition 3.3. *If* $T = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{GL}(2, \mathbb{C})$*, then the* ***trace*** *of* T *is defined by:*

$$\mathrm{Tr}\, T = a + d.$$

Remark 3.4. *If* $T \in \mathrm{Aut}(\hat{\mathbb{C}}) \cong \mathrm{SL}(2, \mathbb{C})/\pm \mathrm{I} \cong \mathrm{PGL}(2, \mathbb{C}$*, then* $\mathrm{Tr}\, T$ *is defined up to multiplication by* ± 1*. Hence in* $\mathrm{Aut}(\hat{\mathbb{C}})$*, only* $\mathrm{Tr}\, T^2$ *is well-defined. Observe also that multiplying a matrix by a scalar* λ *changes the trace by the same factor* λ*, that is the map* $\mathrm{Tr} : \mathrm{GL}(2, \mathbb{C}) \longrightarrow \mathbb{C}$*, defined by* $\mathrm{Tr} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = a + d$*, is linear.*

Remark 3.5. *(i) If we normalize so that* $T \in \mathrm{SL}(2, \mathbb{C})$*, then* $\mathrm{Tr}\, T = \mathrm{Tr}\, T^{-1}$*.*
(ii) $\mathrm{Tr}\, T = \mathrm{Tr}(STS^{-1})$ *for* $S \in \mathrm{SL}(2, \mathbb{C})$*.*

Proof. It is easy to prove (i) by direct calculation. Also check by calculation that, if A and B are matrices in $\mathrm{GL}(2, \mathbb{C})$, then $\mathrm{Tr}(AB) = \mathrm{Tr}(BA)$. Then (ii) follows since

$$\mathrm{Tr}(STS^{-1}) = \mathrm{Tr}(TS^{-1}S) = \mathrm{Tr}(T).$$

□

Corollary 3.6. $\mathrm{Tr}\, T$ *is an invariant under conjugation.*

Lemma 3.7. *Let* $T \in \mathrm{SL}(2, \mathbb{C})$*,* $T \neq Id$*. Then* T *has either 1 or 2 fixed points. The first case occurs* $\iff (\mathrm{Tr}\, T)^2 = 4$*.*

Proof. Let $T = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. Then

$$T(z) = z \iff \frac{az + b}{cz + d} = z \iff cz^2 + (d - a)z - b = 0.$$

So the fixed points of the map T are:

$$z_0^{\pm} = \frac{a - d \pm \sqrt{(d - a)^2 + 4bc}}{2c} = \frac{a - d \pm \sqrt{(\mathrm{Tr}\, T)^2 - 4}}{2c},$$

where the last equality is obtained using the identity $ad - bc = 1$. □

3.1.1 Classification by fixed points and trace.

To analyse the behaviour of transformations in $\mathrm{Aut}(\hat{\mathbb{C}})$, we begin by locating their fixed points. It turns out the trace is sufficient to completely classify the dynamics. We shall analyse the different possibilities one by one. We begin by conjugating so the fixed point(s) are in a particularly easy position to handle, namely 0 and ∞.

Case 1: Parabolic

By Lemma 3.7, T has one fixed point iff $\mathrm{Tr}\, T = \pm 2$ and $T \neq Id$. Call this point z_0. If we conjugate T by a map $S \in \mathrm{Aut}(\hat{\mathbb{C}})$ such that $S(z_0) = \infty$, then $\hat{T} = STS^{-1}$ has a fixed point at ∞, see Example 3.2. So suppose we are in this case.

Let $T = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ such that $T(\infty) = \infty$. So we have that $c = 0, ad = 1$ and hence $T = \begin{pmatrix} a & b \\ 0 & a^{-1} \end{pmatrix}$ and $T(z) = \dfrac{az+b}{a^{-1}} = \alpha z + \beta$ for some α, β. In addition we have the condition $\mathrm{Tr}\, T = a + a^{-1} = \pm 2$ which implies that $a = \pm 1$. Hence $T(z) = z \pm \beta$, that is T is a Euclidean **translation**.

Case 2: Loxodromic

In the generic case, T has 2 fixed points $z_0^{\pm}$. Conjugating T by a map $S \in \mathrm{Aut}(\hat{\mathbb{C}})$ such that $S(z_0^+) = 0$ and $S(z_0^-) = \infty$, that is the map

$$S : z \mapsto \frac{z - z_0^+}{z - z_0^-},$$

we may assume that

$$T = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \quad ad - bc = 1, \quad T(0) = 0,\ T(\infty) = \infty.$$

We deduce that $c = 0, b = 0$ and $d = a^{-1}$ so that

$$T = \begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix}$$

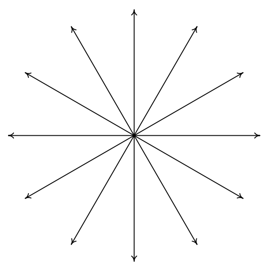

Figure 3.1: Hyperbolic. Expansion $z \mapsto \lambda z, \lambda > 1$.

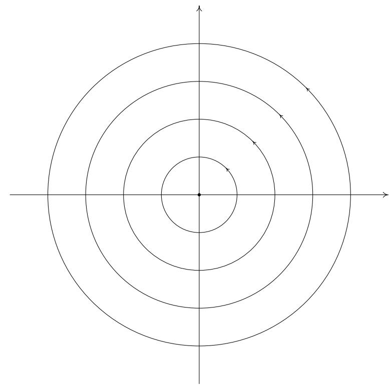

Figure 3.2: Elliptic. Rotation $z \mapsto e^{i\theta}$.

so that

$$T(z) = a^2 z.$$

We also have the condition $\operatorname{Tr} T = a + a^{-1} \neq \pm 2$, equivalently $a \neq \pm 1$. Thus we may write

$$T(z) = \lambda z, \quad \lambda \neq 1.$$

The number λ is called the **multiplier** of T. Note also that $\lambda = T'(0)$.

This can be further subdivided into three subcases:
(a) *Hyperbolic* if $\lambda > 0$;
(b) *Elliptic* if $|\lambda| = 1$;
(c) *Loxodromic* otherwise.

Case 2a: Hyperbolic

This transformation is an expansion (if $\lambda > 1$) or a contraction (if $\lambda < 1$), see Figure 3.1.

Case 2b: Elliptic

This transformation is a rotation as you can see in Figure 3.2.

Case 2c: Loxodromic

This transformation is a combination of the hyperbolic and elliptic cases, so points move along spiral arms. These spirals are exactly what cartographers

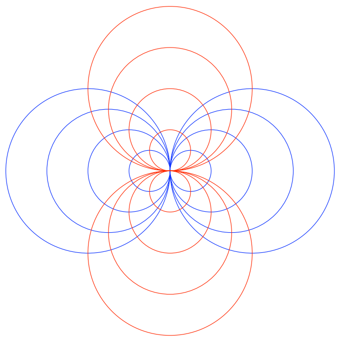

Figure 3.3: Degenerate families of coaxial circles. All the circles in each family are mutually tangent at the common point z_0 and the tangency line of one set is orthogonal to the tangency line of the other.

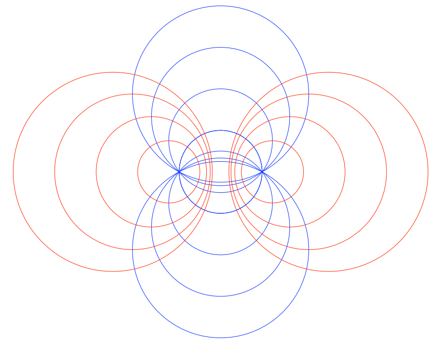

Figure 3.4: Coaxial circles: the generic case. All members of one family of circles pass through the common points $z_0^{\pm}$.

call "loxodromes" derived from the Greek words for "running obliquely".

3.1.2 Coaxial circles

A nice way to understand what happens when the fixed points are not 0 and ∞, is to use **coaxial circles**. These are circles which come in complementary families such that all circles in one family are orthogonal to all circles in the other. There are two cases, the generic one and a degenerate one. In the generic case, all the circles in one family pass through a pair of distinct points, while all the circles in the other family are pairwise disjoint. In the degenerate case, all the circles in both families are tangent at one common point, the tangent line to one family being orthogonal to the tangent line to the other. The two cases are illustrated in Figures 3.3 and 3.4.

Special cases of these two possiblities are illustrated in Figures 3.5 and 3.6. In Figure 3.5 we see the (degenerate) case in which all circles are tangent to the common point ∞. In Figure 3.6 we see the (generic) case in which all circles in one family pass through $0, \infty$ while all those in the other family are pairwise disjoint.

To see that such families of circles exist for general choices of points z_0 or

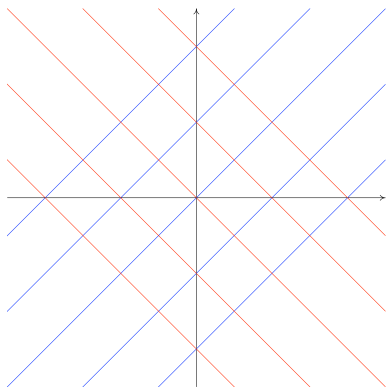

Figure 3.5: Coaxial circles with a common point at ∞.

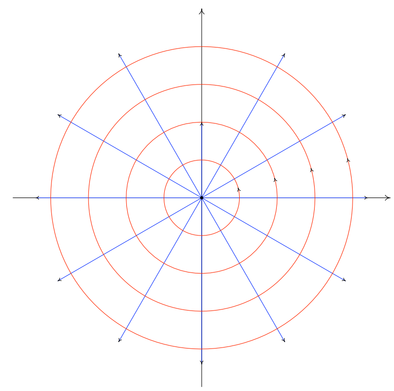

Figure 3.6: Generic coaxial circles: all members of one family all pass through $0, \infty$ while members of the other family are pairwise disjoint.

$z_0^{\pm}$, we apply suitable conjugations to the configurations in Figures 3.5, 3.6, recalling that elements of $\mathrm{Aut}(\hat{\mathbb{C}})$ carry circles to circles and preserve angles. Thus the inverse of the map $z \mapsto \dfrac{1}{z - z_0}$ takes Figure 3.5 to Figure 3.3 and the inverse of the map $z \mapsto \dfrac{z - z_0^+}{z - z_0^-}$ takes Figure 3.6 to Figure 3.4.

3.1.3 Dynamics and Coaxial circles

Now we can return to the dynamics of our various types of transformations. As usual, we first analyse the case when the fixed points are at ∞ and 0.

A parabolic transformation with fixed point at ∞ translates points in a fixed direction. The orbits of points lie along parallel lines through ∞. The orthogonal family of lines is mapped one to another. Likewise a hyperbolic transformation with fixed points at $0, \infty$ has orbits which move inwards or outwards along radial lines from 0. It maps concentric circles centre 0 one into the other. An elliptic with fixed points $0, \infty$ has orbits which move along concentric circles centre 0 and maps radial lines one into the other.

To understand the action of general parabolics, hyperbolics and elliptics, note that conjugation preserves circles and angles, hence coaxial families. The orbits of points under general hyperbolics and parabolics move along a

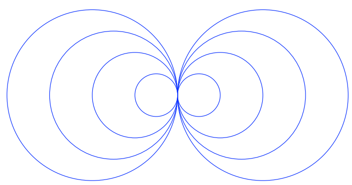

Figure 3.7: Orbit paths for a parabolic with fixed point z_0.

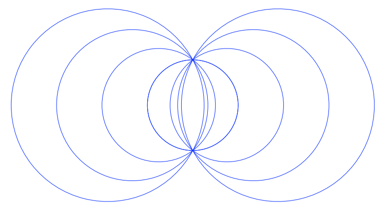

Figure 3.8: Orbit paths for a hyperbolic with fixed point $z_0^{\pm}$.

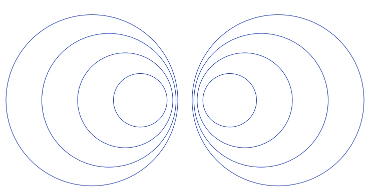

Figure 3.9: Orbit paths for an elliptic. The circles of the complementary family all pass through the two fixed points.

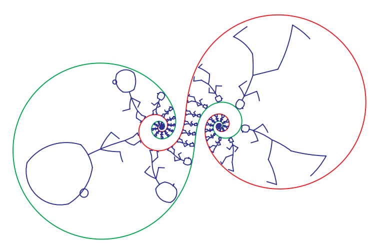

Figure 3.10: Orbit paths for a loxodromic. Picture from *Indra's Pearls* [2].

family of coaxial circles through the fixed point(s) and map members of the orthogonal family one into the other, as shown in Figures 3.7 – 3.10.
Exercise: Describe the movement of an elliptic in a similar way.

3.1.4 Relation of multiplier to trace

Consider a *loxodromic* transformation $T : z \mapsto \lambda z, \lambda \in \mathbb{C}$. Recall that λ is called the **multiplier** of T. Write

$$T = \pm \begin{pmatrix} \sqrt{\lambda} & 0 \\ 0 & \sqrt{\lambda^{-1}} \end{pmatrix}, \ \det T = 1, \ \operatorname{Tr} T = \sqrt{\lambda} + \sqrt{\lambda^{-1}}.$$

Set $l = \log \lambda$ (defined only modulo $2\pi i$). Then $e^l = \lambda$, so

$$\operatorname{Tr} T = e^{\frac{l}{2}} + e^{-\frac{l}{2}} = 2\cosh(l/2).$$

(Note that $l/2$ is defined only modulo πi and that $\cosh(x + i\pi) = -\cosh x$, so there is the same ambiguity as in the trace.) Writing $l/2 = x + iy$ we note that $\cosh(x+iy) = \cosh x \cos y - i \sinh x \sin y$. In particular, $\cosh(x+iy) = 1$ if and only if $x = 0$ and $y = in\pi$. Thus

$$\begin{aligned} \operatorname{Tr} T &= \pm 2 \iff \cosh(l/2) = \pm 1 \iff \\ x &= 0 \text{ and } y = n\pi i \iff l = 2n\pi i \iff \lambda = 1 \iff T = \mathrm{id}. \end{aligned}$$

Now note that:

$$\lambda > 0, \lambda \neq 1 \iff l \in \mathbb{R}^* + 2ni\pi \iff \operatorname{Tr} T \in \mathbb{R}, |\operatorname{Tr} T| > 2.$$

$$|\lambda| = 1, \lambda \neq 1 \iff l \in i\mathbb{R}^* \iff \operatorname{Tr} T \in (-2, 2)$$

where $\mathbb{R}^* = \mathbb{R} \setminus \{0\}$. We have proved the following:

Proposition 3.8. *Let T be a transformation in $\mathrm{SL}(2, \mathbb{C})$. Then: (i) T is parabolic (or is the identity map) $\iff \operatorname{Tr} T = \pm 2$;*
(ii) T is hyperbolic $\iff \operatorname{Tr} T \in \mathbb{R}, |\operatorname{Tr} T| > 2$;
(iii) T is elliptic $\iff \operatorname{Tr} T \in (-2, 2)$;
(iv) T is loxodromic $\iff \operatorname{Tr} T \notin \mathbb{R}$.

Lemma 3.9. *The multiplier and its inverse λ and λ^{-1} are the derivatives of T at its fixed points. If $\lambda > 1$, then it is the derivative of T at the repelling fixed point, while if $\lambda < 1$, then it is the derivative of T at the attracting fixed point.*

Proof. This is easy to check if $T(z) = \lambda z$ so that its fixed points are $0, \infty$. We may assume without loss that $\lambda > 1$ so that 0 is the repelling point In general, suppose T has attracting and repelling fixed points z^+, z^- respectively. Let $S \in SL(2,\mathbb{C})$ carry $\infty \mapsto z^+$ and $0 \mapsto z^-$. Then $T = S\hat{T}S^{-1}$ where $\hat{T}$ has attracting and repelling fixed points at $\infty, 0$ respectively. Moreover by the chain rule $T'(z^-) = S'(\hat{T}S^{-1}(z^-))\hat{T}'(S^{-1}(z^-))(S^{-1})'(z^-))$. Now $\hat{T}'(S^{-1}(z^-)) = \hat{T}'(0) = \lambda$ and since $SS^{-1} = \mathrm{id}$ we have $S'(S^{-1}(z^-))(S^{-1})'(z^-) = 1$. This gives the result. □

3.1.5 Isometric circles

A concept which is sometimes useful is that of the **isometric circle** of $T \in \operatorname{Aut}\hat{\mathbb{C}}$. This is a circle on which it neither expands not contracts, ie on which $|T'(z)| = 1$. To compute this:

If $T = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, then $T'(z) = \dfrac{1}{(cz+d)^2}$. So

$$|T'(z)| = 1 \iff |cz+d| = 1.$$

This is a circle with centre $-d/c$ and radius $1/c$. Inside that circle we have that $|T'(z)| > 1$, while outside we have that $|T'(z)| < 1$.

Exercise 3.10. *If I_T is the isometric circle of T, show that $T(I_T) = I_{T^{-1}}$ and that $I_{ST} \neq S(I_T)$.*

3.2 Isometries of $\mathbb{H}$ and $\mathbb{D}$

Now we apply the above results to the classification of the orientation preserving isometries of $\mathbb{H}$ and $\mathbb{D}$.

Proposition 3.11. *If $T \in \operatorname{Aut}\mathbb{H}$ or $T \in \operatorname{Aut}\mathbb{D}$, then $\operatorname{Tr} T \in \mathbb{R}$ (and $\det T > 0$).*

Proof. If $T \in \operatorname{Aut}\mathbb{H} \cong \mathrm{SL}(2,\mathbb{R})$, this is obvious. But it is true also for $T \in \operatorname{Aut}\mathbb{D}$, since if $C : \mathbb{H} \longrightarrow \mathbb{D}$ is the Cayley transformation, then $\operatorname{Aut}(\mathbb{D}) = C \operatorname{Aut}\mathbb{H} C^{-1}$ and the trace and the determinant are invariant under conjugation. □

Now consider $\operatorname{Aut}\mathbb{H} = \mathrm{SL}(2,\mathbb{R})$. Let $T = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, $\det T = 1$, $\operatorname{Tr} T \in \mathbb{R}$.

The fixed points of $T \in \operatorname{Aut}\mathbb{H}$ are the solutions $z_0^{\pm}$ of the equation $z = \dfrac{az+b}{cz+d}$ and they can be calculated, as we have proved previously, by the formula

$$z_0^{\pm} = \frac{a - d \pm \sqrt{(\operatorname{Tr} T)^2 - 4}}{2c}.$$

It follows from Proposition 3.8 and the fact that the coefficients a, b, c, d are real valued, that there are three possible cases to consider:[1]

(i) T is a **parabolic** transformation if $\operatorname{Tr} T = \pm 2$ and $T \neq Id$. Then T has one fixed point in $\hat{\mathbb{R}}$.

(ii) T is a **hyperbolic** transformation if $|\operatorname{Tr} T| > 2$. Then T has two fixed points in $\hat{\mathbb{R}}$.

(iii) T is an **elliptic** transformation if $|\operatorname{Tr} T| < 2$. Then T has two complex conjugate fixed points, so exactly one of these points is in $\mathbb{H}$.

If $T \in \operatorname{Aut}\mathbb{D}$, then conjugating by the Cayley transform we see that the three cases are almost the same. The only difference is that parabolics and hyperbolics have their fixed points on $\partial\mathbb{D}$ while the elliptics have exactly one of their two fixed points inside $\mathbb{D}$, and one outside. (Why?) We shall look in detail at each of these three cases in turn.

[1] The terminology can be explained by looking at the linear action of the corresponding matrices on $\mathbb{R}^2$. A matrix in $\mathrm{SL}(2,\mathbb{R})$ is hyperbolic if and only if it is diagonalizable over $\mathbb{R}$, or conjugate in $\mathrm{SL}(2,\mathbb{R})$ to $\begin{pmatrix} \lambda & 0 \\ 0 & \frac{1}{\lambda} \end{pmatrix}$. It is elliptic iff it is conjugate in $\mathrm{SL}(2,\mathbb{R})$ to $\begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$, and parabolic iff it is conjugate to $\begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}$. Under the usual linear action of $\mathrm{SL}(2,\mathbb{R})$ on $\mathbb{R}^2$, the invariant curves for the first two canonical forms are respectively rectangular hyperbolae $xy = constant$ and circles centre 0. Applying a transformation in $\mathrm{SL}(2,\mathbb{R})$ carries these to hyperbolae and ellipses, hence the terminology. You can think of the parabolics, whose invariant curves are lines, as an intermediate case.

3.2.1 Parabolic transformations

This is the case $\operatorname{Tr} T = \pm 2$ and $T \neq Id$. Such transformations have exactly one fixed point $\operatorname{Fix} T$ on $\partial\mathbb{H}$ or $\partial\mathbb{D}$.

First suppose $T \in \operatorname{Aut}\mathbb{H}$ and conjugate by an element $S \in \mathrm{SL}(2,\mathbb{R})$ so that $\operatorname{Fix} T$ moves to ∞. To do this, we can use the transformation

$$S : z \mapsto -\frac{1}{z-\xi}, \quad \xi = \operatorname{Fix} T.$$

If

$$T \in \mathrm{SL}(2,\mathbb{R}),\ T(\infty) = \infty,\ T = \begin{pmatrix} a & b \\ c & d \end{pmatrix},\ \det T = 1,$$

then, as we proved in Section 3.1.1,

$$T(z) = z + \alpha, \quad \alpha \in \mathbb{R}.$$

This is a Euclidean **translation**. It fixes the horocycles $\{z \in \mathbb{H} : \Im z > t > 0\}$, that is the horizontal lines shown in Figure 3.11, and maps the orthogonal family of vertical geodesics to itself. Note that the induced metric $ds = dx/Imz$ on a horocycle is Euclidean since $y = \Im z$ is constant.

Conjugating back to the case where the fixed point is $\xi \in \mathbb{R}$ using S^{-1} gives Figure 3.12. The horocycles are tangent to $\mathbb{R}$ at ξ, because the horizontal horocycles are tangent to $\partial\mathbb{H}$ at ∞. The orthogonal family of circles hit $\partial\mathbb{H}$ orthogonally and so are geodesics in $\mathbb{H}$. T maps one of these geodesics into another.

Exercise 3.12. *Show that any parabolic transformation which fixes* 0 *belongs to the family* $T_c(z) = \dfrac{z}{cz+1}$, $c \in \mathbb{R} \setminus \{0\}$.

HINT: The transformation $J : z \mapsto -1/z$ conjugates $\begin{pmatrix} 1 & 0 \\ c & 1 \end{pmatrix}$ to $\begin{pmatrix} 1 & -c \\ 0 & 1 \end{pmatrix}$.

□

3.2.2 Hyperbolic transformations

This is the case $|\operatorname{Tr} T| > 2$. Such transformations have two fixed points, both of which are on $\mathbb{R} \cup \infty$.

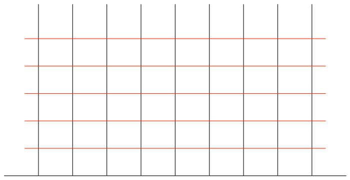

Figure 3.11: A parabolic with fixed point at ∞. Points move along the horizontal lines and map the vertical lines one into the other.

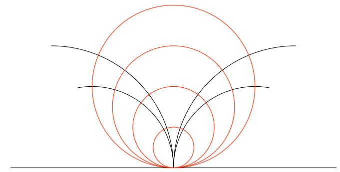

Figure 3.12: Action of a parabolic in general position.

Suppose that $T \in \mathrm{Aut}\,\mathbb{H}$ and conjugate T by an element $S \in \mathrm{SL}(2,\mathbb{R})$ so that the fixed points are 0 and ∞. This can be done using the transformation

$$S : z \mapsto -\frac{z-\xi^+}{z-\xi^-}, \quad \xi^\pm = \mathrm{Fix}\,T.$$

Then as in Section 3.1.1, $T(z) = \lambda z$ where $\lambda > 0$.

The picture is shown in Figure 3.13. T maps the radial lines to themselves and the orthogonal family of semicircular geodesics to themselves. Notice that the radial lines are the lines equidistant from the imaginary axis, which is the line joining the fixed points of T. Conjugating back to the general case, we obtain Figure 3.14.

Axis and Translation length

The geodesic joining the fixed points of T is called the **axis** of T and is denoted $\mathrm{Ax}\,T$. The **translation length** is the distance a point is moved by T along its axis, ie $d_{\mathbb{H}}(P, T(P))$. It turns out that this is independent of P.

To see this, it is sufficient to compute $d_{\mathbb{H}}(P, T(P))$ when $T(z) = \lambda z$ and $P \in i\mathbb{R}$. (Why?) So suppose that $P = it$. Then $d_{\mathbb{H}}(P, T(P)) = d_{\mathbb{H}}(it, \lambda it) = \log \lambda = l$ say. Note that $l = l(T)$ is independent of P. By the results of Section 3.1.4, we see that if $l = \log \lambda$, then $\mathrm{Tr}\,T = 2\cosh l(T)/2$. Notice that both sides are invariant under conjugation. This gives a direct relationship between the distance moved and the trace, and ascribes a geometric meaning to $\mathrm{Tr}\,T$.

We can describe T as a **hyperbolic translation** along its axis.

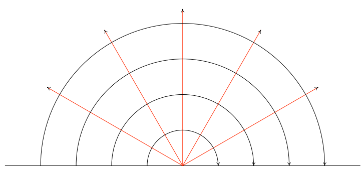

Figure 3.13: A hyperbolic with fixed points at $0, \infty$. Points move along the equidistant curves and map the orthogonal family of geodesics one into the other.

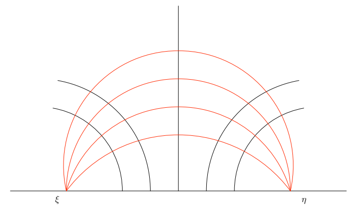

Figure 3.14: A hyperbolic in general position. Points move along the equidistant curves and map the orthogonal geodesics one into the other.

Distance moved by other points

Now suppose that $P \notin \operatorname{Ax} T$. What can we say about $d_{\mathbb{H}}(P, T(P))$? As usual, put T in standard position so $T(z) = \lambda z$, $\lambda > 0$, and let $P = re^{i\theta}$. Then $T(P) = \lambda re^{i\theta}$. So

$$d_{\mathbb{H}}\left(P, T(P)\right) = d_{\mathbb{H}}(re^{i\theta}, \lambda re^{i\theta}) = \rho \quad \text{where} \quad \cosh\rho = 1 + \frac{|1-\lambda|^2}{2\lambda\sin^2\theta}.$$

To interpret this formula, suppose that d is the distance of P from $i\mathbb{R}$. The angle of parallelism formula gives $\cosh d = 1/\sin\theta$. (Notice this depends only on θ and not on r.) The formula relating the multiplier λ to $\operatorname{Tr} T = 2\cosh l/2$ gives

$$\frac{|1-\lambda|^2}{\lambda} = 4\sinh^2 l/2.$$

Hence

$$\cosh\rho = 1 + 2\sinh^2\rho/2 = 1 + \frac{2\sinh^2 l/2}{\sin^2\theta}$$

so that

$$\sinh\rho/2 = \sinh l/2\cosh d.$$

In particular, $\rho > l$ and P moved further the further it is from the axis of T. (We remark that this formula is a special case of Theorem 2.23.)

Attracting and repelling fixed points

QUESTION: How do we discover which fixed point is attracting and which is repelling? There are several ways to answer this:

Method 1: Compute the image of some point which is not the fixed point.

Method 2: Compute the derivative $T'(z) = (cz + d)^{-2}$ at the fixed points. The attractive fixed point z_+ is the one for which $|T'(z_+)| < 1$.

Method 3: One can devise a formula but has to take care of the square roots. This is done in [2] p. 84

Example 3.13. *Let* $T = \begin{pmatrix} 2 & 3 \\ 3 & 5 \end{pmatrix}$.

Note first that $\det T = 1$. Then T is hyperbolic because $\operatorname{Tr} T \in \mathbb{R}$ and $|\operatorname{Tr} T| > 2$. The translation length is given by l where $\operatorname{Tr} T = 2\cosh \frac{l}{2} = 7$, so $l = 2\text{arccosh}\frac{7}{2}$.
The fixed points are

$$z^{\pm} = \frac{a - d \pm \sqrt{(\operatorname{Tr} T)^2 - 4}}{2c} = \frac{-3 \pm \sqrt{45}}{6} = \frac{-1 \pm \sqrt{5}}{2}.$$

Let us find which fixed point is the attracting one using both Method 1 and Method 2 above:
Method 1:
Calculate $T(0) = \frac{3}{5} > 0$. Thus T moves points away from $z^- = \frac{-1-\sqrt{5}}{2}$ and towards $z^+ = \frac{-1+\sqrt{5}}{2}$, from which it follows that the attracting fixed point is z^+.
Method 2:

$$T'(z^+) = T'\left(\frac{-1+\sqrt{5}}{2}\right) = \left(\frac{7}{2} + \frac{3\sqrt{5}}{2}\right)^{-2} \Rightarrow \left|T'\left(\frac{-1+\sqrt{5}}{2}\right)\right| < 1.$$

Thus z^+ is the attracting fixed point. □

Example 3.14. *The family of transformations* $S_t = \begin{pmatrix} \cosh t & \sinh t \\ \sinh t & \cosh t \end{pmatrix}$, $t > 0$.

We claim this is the family of all hyperbolic transformations in $\operatorname{Aut}\mathbb{D}$ with fixed points in 1 and -1. First check easily that $S_t \in \operatorname{Aut}\mathbb{D}$ and that $\operatorname{Tr} S_t = 2\cosh t$ so that S_t is hyperbolic. In particular, S_t translates points on its axis by a distance $2t$.

You can either compute the fixed points of S_t directly, or argue as follows. The Cayley transformation $C : \mathbb{H} \longrightarrow \mathbb{D}$ sends the points $0, \infty \longrightarrow -1, +1$. Therefore we can conjugate this family to the family

$$T_\lambda : z \mapsto \lambda z,$$

where the fixed points are 0 and ∞ and where

$$\operatorname{Tr} T_\lambda = 2\cosh l/2 \quad \text{with} \quad l = \log \lambda.$$

Comparing traces we find $CT_\lambda C^{-1} = S_t$ with $t = \log \lambda/2$.□

3.2.3 Elliptic transformations

This is the case $|\operatorname{Tr} T| < 2$. Such transformations have two conjugate fixed points, exactly of which is in $\mathbb{H}$ (or $\mathbb{D}$).

Suppose that $T \in \operatorname{Aut}\mathbb{H}$ and conjugate by an element $S \in \mathrm{SL}(2, \mathbb{R})$ so that the fixed points are $\pm i$. (Why is this possible?) We calculate easily that

$$T(i) = i \Rightarrow T = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}.$$

In fact,

$$i = \frac{ai + b}{ci + d} \Rightarrow a = d, b = -c, a^2 + b^2 = 1.$$

Such a map preserve hyperbolic circles centered at i, as you can see in Figure 3.15. The dotted circles in the lower half plane are outside $\mathbb{H}$ and so not part of the action in our model.

The action carries the geodesics emanating from i (the black lines in Figure 3.15) one into the other. The **angle of rotation** is the angle between one such geodesic and its image under T. To calculate this angle, it is enought to compute the derivative at i, that is

$$T'(i) = \frac{1}{(-i\sin\theta + \cos\theta)^2} = e^{2i\theta}.$$

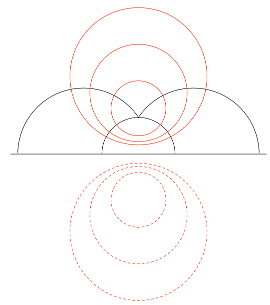

Figure 3.15: Action of an elliptic with fixed point at i. Points move around the hyperbolic circles centre i and the orthogonal geodesics are mapped one into the other.

So T is a **hyperbolic rotation** by $\psi = 2\theta$ about i.

We can relate the rotation angle to $\operatorname{Tr} T$ by observing that $\operatorname{Tr} T = 2\cos\theta$. Since both $\operatorname{Tr} T$ and the rotation angle $\psi(T)$ are clearly invariant under conjugation (why?) we find

$$\operatorname{Tr} T = 2\cos\psi(T)/2.$$

This should be compared with the formula for translation length of a hyperbolic:

$$\operatorname{Tr} T = 2\cosh l(T)/2.$$

Example 3.15. *Conjugating* $T = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$ *to* $\mathbb{D}$ *using the Cayley transformation* $C : \mathbb{H} \longrightarrow \mathbb{D}$, *we obtain the map*

$$\hat{T}(z) = e^{2i\theta} z.$$

PROOF: We calculate:

$$\hat{T} = CTC^{-1} = \begin{pmatrix} 1 & -i \\ 1 & i \end{pmatrix} \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} i & i \\ -1 & 1 \end{pmatrix} \cdot \frac{1}{(1+i)^2}$$

$$= \begin{pmatrix} e^{i\theta} & -ie^{i\theta} \\ e^{-i\theta} & ie^{-i\theta} \end{pmatrix} \begin{pmatrix} i & i \\ -1 & 1 \end{pmatrix} \cdot \frac{1}{2i} = \begin{pmatrix} 2ie^{i\theta} & 0 \\ 0 & 2ie^{-i\theta} \end{pmatrix} \cdot \frac{1}{2i}.$$

Hence $\hat{T}(z) = e^{2i\theta}z$, confirming that the rotation is by angle 2θ. We can also see directly that $\operatorname{Tr}\hat{T} = 2\cos\theta$.
The action of an elliptic transformation in general position is similar.

Remark 3.16. *In 3-dimensions the two formulae for elliptic and hyperbolic transformations are part of one complex formula for complex translation distance which can be* ***defined*** *to be $l(T) = \log\lambda/2$ where λ is the multiplier. To see how this works in the elliptic case, if the multiplier is λ then*

$$\log\lambda = \log(e^{2i\theta}) = 2i\theta$$

and that

$$\operatorname{Tr} T = 2\cosh\frac{2i\theta}{2} = 2\cos\theta.$$

The ambiguity of πi in defining $\log\lambda/2$ *turns into the ambiguity of* ± 1 *in defining the trace.*

3.2.4 Orientation Reversing Isometries

To understand the orientation reversing isometries of $\mathbb{H}$ and $\mathbb{D}$, it is enough to find a single example. This is because if S, S' are orientation reversing isometries then $S'S^{-1}$ (or indeed $S'S$) preserves orientation.

Example 3.17. *The map $T : z \mapsto -\bar{z}$ is a reflection in $i\mathbb{R}^+$ and reverses orientation. Obviously it is an isometry of $\mathbb{H}$.*

By conjugation, we get an orientation reversing isometry with fixed axis any line $\mathcal{L}$ in $\mathbb{H}$. We can use the Cayley transformation to find analogous isometries in $\mathbb{D}$. This gives:

Proposition 3.18. *Given any hyperbolic line $\mathcal{L}$, there exists a unique orientation reversing isometry $\sigma_{\mathcal{L}}$ such that $\sigma_{\mathcal{L}}|_{\mathcal{L}} = id, \sigma_{\mathcal{L}}^2 = id, d_{\mathbb{H}}(x, \mathcal{L}) = d_{\mathbb{H}}(\sigma_{\mathcal{L}}, \mathcal{L})$.*

Geometrically $\sigma_{\mathcal{L}}$ is the **inversion in the circle** $\mathcal{L}$. If $\mathcal{L}$ has Euclidean radius ρ and centre O, then inversion in $\mathcal{L}$ is the map which sends radial lines to themselves sending $P \mapsto P'$ where $OP \cdot OP' = \rho^2$. (See Example Sheet 5.)

Exercise 3.19. *We can investigate a general orientation reversing isometry in* $\mathrm{Isom}^-\mathbb{H}$ *computationally as follows. Why can any such map be written in the form* $T : z \mapsto \dfrac{a\bar{z}+b}{c\bar{z}+d}$*? We calculate*

$$\Im(T(z)) = \frac{bc-ad}{|cz+d|^2}.$$

Hence in order for T to map $\mathbb{H}$ to itself, we need $ad - bc = -1$.

The equation which defines the fixed points of T is

$$cz\bar{z} + dz - a\bar{z} - b = 0.$$

So, if we set $z = x + iy$, then:

$$(a+d)y = 0 \quad \text{and} \quad c(x^2+y^2) + (d-a)x - b = 0.$$

From the first equation we get $a = -d$ or $y = 0$. If $y = 0$ then $x = \dfrac{a-d \pm \sqrt{(\mathrm{Tr}\,T)^2+4}}{2c}$, so T has two fixed points on $\hat{\mathbb{R}}$ (and no other fixed points). If $a = -d$ then $c(x^2+y^2) + (d-a)x - b = 0$ is the equation of a circle with centre on $\mathbb{R}$. In other words, T has a fixed geodesic in $\mathbb{H}$.

The map $T(z) = -\lambda\bar{z}$ is a product of the reflection $z \mapsto \bar{z}$ and an homothety. How does this fit with the computations above?

3.2.5 Isometries as product of reflections

Recall that, in Euclidean geometry, any translation (resp. any rotation) is a product of two reflections in parallel (resp. intersecting) lines. In hyperbolic geometry, we have a similar fact, but now there are 3 cases:
(i) reflection in two disjoint (ultra-parallel) lines gives a hyperbolic transformation;
(ii) reflection in two lines meeting at ∞ gives a parabolic transformation;
(iii) reflection in two intersecting lines gives an elliptic transformation.

We will analyze the first case in detail: the other two are similar.

Proposition 3.20. *Let $\mathcal{L}$ and $\mathcal{M}$ two disjoint lines and let $\sigma_{\mathcal{L}}$ and $\sigma_{\mathcal{M}}$ be the reflections in $\mathcal{L}$ and $\mathcal{M}$ respectively. Then $\sigma_{\mathcal{M}}\sigma_{\mathcal{L}}$ is the hyperbolic* *transformation whose axis N is the common perpendicular of $\mathcal{L}$ and $\mathcal{M}$. It has translation distance $2d$ where d is the distance between $\mathcal{L}$ and $\mathcal{M}$ and moves in the direction from $\mathcal{L}$ towards $\mathcal{M}$.*

Proof. Observe that $\sigma_{\mathcal{L}}$ and $\sigma_{\mathcal{M}}$ preserves N. If P be the point $N \cap \mathcal{L}$, then $\sigma_{\mathcal{M}}\sigma_{\mathcal{L}}(P)$ is at distance $2d$ from P on the other side of $\sigma_{\mathcal{M}}$. Why is this enough? □

The following consequence gives the flavour of properties we shall be studying in the next chapter.

Corollary 3.21. *Any non-abelian subgroup G of* $\mathrm{SL}(2, \mathbb{R})$ *contains hyperbolic elements.*

Proof. Let the subgroup be H and suppose first that $T \in H$ is elliptic with fixed point v. Pick $S \in H$ such that $ST \neq TS$. Then $S(v) \neq v$. (Why?)

Observe that if T is anticlockwise rotation of 2θ about v, then STS^{-1} is anticlockwise rotation by 2θ round $S(v)$. So $ST^{-1}S^{-1}$ is anticlockwise rotation by -2θ round $S(v)$.

Let $\mathcal{N}$ be the line joining v to $S(v)$ and let $\mathcal{L}$ and $\mathcal{M}$ be lines through v and $S(v)$ such that the anticlockwise angle from $\mathcal{N}$ to both $\mathcal{L}$ and $\mathcal{M}$ is 2θ. Then $\mathcal{L}$ and $\mathcal{M}$ are disjoint (Why?) and furthermore, we see that $T = \sigma_{\mathcal{L}}\sigma_{\mathcal{N}}$ and $ST^{-1}S^{-1} = \sigma_{\mathcal{N}}\sigma_{\mathcal{M}}$. It follows that $TST^{-1}S^{-1} = \sigma_{\mathcal{L}}\sigma_{\mathcal{M}}$. By Proposition 3.20, $\sigma_{\mathcal{L}}\sigma_{\mathcal{M}}$ is a hyperbolic translation along the common perpendicular of $\mathcal{L}$ and $\mathcal{M}$. The case in which T is parabolic is a similar exercise. □

Chapter 4

Groups of isometries

4.1 Subgroups of $\mathrm{Isom}^+(\mathbb{H})$

In this chapter we start to look at discrete groups of isometries of $\mathbb{H}$. This encompasses the symmetry groups of tessellations and, what turns out to be very closely related, the fundamental groups of surfaces. After defining a discrete group, we begin by exploring the cyclic and abelian groups.

Definition 4.1. *A group G is a topological group if G is a topological space for which the multiplication maps $G \times G \longrightarrow G, (x, y \mapsto xy)$ and the inverse map $G \longrightarrow G, x \mapsto x^{-1}$ are continuous.*

Lemma 4.2. *The groups* $\mathrm{SL}(2,\mathbb{R})$ *and* $\mathrm{PSL}(2,\mathbb{R})$ *are topological groups.*

Proof. We define the topology on $\mathrm{SL}(2,\mathbb{R})$ by identifiying it with the subset $X \subset \mathbb{R}^4$ defined by

$$\{(a, b, c, d) \in \mathbb{R}^4 : ad - bc = 1\}.$$

If we define the map $\delta : X \longrightarrow X$ mapping $(a, b, c, d) \longrightarrow (-a, -b, -c, -d)$, then δ is a homeomorphism and δ together with the identity forms a cyclic group of order 2 acting on X. We topologize $\mathrm{PSL}(2,\mathbb{R})$ as the quotient space.

It is clear that multiplication and inverse are continuous in both $\mathrm{SL}(2,\mathbb{R})$ and $\mathrm{PSL}(2,\mathbb{R})$. □

We shall be mainly interested in subgroups of $\mathrm{SL}(2,\mathbb{R})$ which satisfy one of two closely related conditions, that of being **discrete** in $\mathrm{SL}(2,\mathbb{R})$ and of acting **properly discontinuously** on $\mathbb{H}$. It is an important property of subgroups of $\mathrm{SL}(2,\mathbb{R})$ that these two conditions are in fact equivalent. We shall prove this as Theorem 4.20. in section 4.2. For the moment, we study the two definitions and some variants and look at a collection of examples.

Definition 4.3. *A subgroup $G \subset \mathrm{SL}(2,\mathbb{R})$ is **discrete** if G has no accumulation points in $\mathrm{SL}(2,\mathbb{R})$.*

Definition 4.4. *The group G acts **properly discontinuously** on $\mathbb{H}$ if for all compact subsets $K \subset \mathbb{H}$, then $gK \cap K = \varnothing$ for all but finitely many $g \in G$.*

The condition of being discrete can also be formulated in slightly different ways.

Lemma 4.5. *The following conditions on subgroup $G \subset \mathrm{SL}(2,\mathbb{R})$ are equivalent:*
(i) There are no accumulation points in G;
(ii) G has no accumulation points in $\mathrm{SL}(2,\mathbb{R})$;
(iii) The identity is an isolated point of G.

Remark 4.6. *Condition (ii) is on the face of it somewhat stronger than (i). In fact for general metric spaces $X \subset Y$, the conditions (i) and (ii) are not equivalent. For example, suppose $X = \{\frac{1}{n}\} \subset [0,1] = Y$. Then X has no accumulation points in itself, so (i) holds, but it does have an accumulation point $0 \in Y$, so (ii) fails. The proof which follows shows that if X, Y are topological groups, then the two conditions are the same.*

Proof. (i) implies (ii): Suppose that (ii) fails so that $g_n \longrightarrow h$ for some $h \in \mathrm{SL}(2,\mathbb{R})$. Then $g_n g_{n+1}^{-1} \longrightarrow hh^{-1} = \mathrm{id}$. Since G is a subgroup, id ∈ G and hence (i) fails.

(ii) implies (iii) since if the identity is not an isolated point, then clearly (ii) fails.

(iii) implies (i): Choose a neighbourhood id ∈ U such that $G \cap U = \{\mathrm{id}\}$. Then if $g \in G$, we have $G \cap gU = \{g\}$ which implies (i). □

The condition of acting properly discontinuously depends on properties of the space being acting on. We shall see below that any discrete subgroup of $\mathrm{SL}(2,\mathbb{R})$ acts properly discontinuously on $\mathbb{H}$. However:

Example 4.7. *The group $\mathrm{SL}(2,\mathbb{Z})$ of 2×2 matrices with integer coefficients and determinant 1 is obviously a discrete subgroup of $\mathrm{SL}(2,\mathbb{R})$. As we shall see below, any discrete subgroup of $\mathrm{SL}(2,\mathbb{R})$ acts properly discontinuously on*

$\mathbb{H}$. However $\mathrm{SL}(2,\mathbb{Z})$ does not act properly discontinuously on $\mathbb{R}\cup\infty$. To see this, let $I=[0,1]$ and consider $g_n = \begin{pmatrix} 1 & -1 \\ n+1 & -n \end{pmatrix} \in \mathrm{SL}(2,\mathbb{Z})$. Then $g_n O = 1/n \in I$ for $n>1$, and so $g_n I \cap I \neq \emptyset$. Since I is compact this proves the claim.

There are also a number of variants on the definition of acting properly discontinuously. We have chosen the one which works best in our proofs. The next lemma gives some examples. For further discussion, see [14] Section 3.5, also Examples Sheet 6.

Lemma 4.8. *Let G be a subgroup of $\mathrm{SL}(2,\mathbb{R})$ The following conditions are equivalent:*
(i) G does not act properly discontinuously on $\mathbb{H}$.
(ii) Some G orbit in $\mathbb{H}$ has accumulation points in $\mathbb{H}$.
(iii) All G orbits in $\mathbb{H}$ have accumulation points in $\mathbb{H}$, with the possible exception of one orbit which consists of a single point fixed by all elements in G.

Proof. (i) implies (ii): If G is not PD, then there exists a compact set K with $g_n K \cap K \neq \emptyset$ for infinitely many distinct $g_n \in G$. Thus there are points $z_n \in K$ with $g_n z_n \in K$. Since K is compact we can assume by passing to a subsequence that $z_n \longrightarrow w \in K$.

Now K is bounded and so is contained in the closed ball $\overline{B_R(w)}$ for some $R>0$. Since $z_n \longrightarrow w \in K$, we have $d(z_n, w) < 1$ for large enough n and so also $d(g_n z_n, g_n w) < 1$. Since $g_n z_n \in K$ it follows that $g_n w$ lie in the closed ball $\overline{B_{R+1}(w)}$ which is compact. Thus we can extract a convergent subsequence, showing that the G orbit of w has an accumulation point in $\mathbb{H}$.

(ii) implies (iii): Suppose the orbit of z_0 has an accumulation point, so there exist distinct points $g_n z_0 \longrightarrow w_0 \in \mathbb{H}$. Let $z \in \mathbb{H}$ and consider the points $g_n z$. We have

$$d(g_n z, z) \leq d(g_n z, g_n z_0) + d(g_n z_0, w_0) + d(w_0, z) = d(z, z_0) + d(g_n z_0, w_0)) + d(w_0, z).$$

The right hand side of this expression is bounded independently of n, hence the points $g_n z$ are all contained in a closed ball centre z. If infinitely many of these points are distinct, it follows that the orbit has an accumulation point.

If $g_n z$ coincide for infinitely many distinct n, then we must have $h_n z = z$ for infinitely many distinct elements $h_n \in G$. This means that z must be a common fixed point of the h_n which are all elliptic. We claim that the orbit of any other point $w \neq z$ has accumulation points. For since the h_n are all distinct, they all rotate through different angles, and hence there must be infinitely many distinct points $h_n w$ which all lie on a circle centre z. This circle is compact and so the orbit of w has an accumulation point.

(iii) implies (i): Pick an orbit Gz say with an accumulation point, so that $g_n z \longrightarrow w$ for some $g_n \in G$ and $w \in \mathbb{H}$. Choose $R > 0$ so that the closed ball $\overline{B_R(w)}$ contains z. Clearly $\overline{B_R(w)}$ contains infinitely many points $g_n z$, so $g_n \overline{B_R(w)} \cap \overline{B_R(w)} \neq \emptyset$ for infinitely many g_n. Since $\overline{B_R(w)}$ is compact the result follows. □

We can now prove half of Theorem 4.20 referred to above:

Corollary 4.9. *If G acts properly discontinuously on $\mathbb{H}$ then G is discrete in* $\mathrm{SL}(2,\mathbb{R})$.

Proof. If G is not discrete then by Lemma 4.5 (iii) we can find a sequence $g_n \in G$ such that $g_n \longrightarrow id$. Thus $g_n z \longrightarrow z \ \forall \ z \in \mathbb{H}$. We want to show there is a G-orbit with an accumulation point, so we have to show that for some choice of $z = z_0$, the points $g_n z_0$ are all distinct.

We do this by choosing a point $z_0 \in \mathbb{H}$ which is not a fixed point of any $g_m^{-1} g_n, n, m \in \mathbb{N}$. Then the points $g_n z_0$ are all distinct and $g_n z_0 \longrightarrow z_0$. So z_0 is an accumulation point of the G-orbit of z_0, and so by Lemma 4.8, G does not act properly discontinuously. □

Exercise 4.10. *Show that both the properties of being discrete and of acting properly discontinuously on $\mathbb{H}$ are conjugation invariant, ie if G has one of these properties, then so does the group $TGT^{-1} = \{TgT^{-1} : g \in G\}$, where $T \in \mathrm{SL}(2,\mathbb{R})$.*

Now we collect some examples of subgroups of $\mathrm{Isom}^+ \mathbb{H}$.

Example 4.11. *If $T \in \mathrm{Isom}^+ \mathbb{H}$ then $G =< T >= \{T^n : n \in \mathbb{Z}\}$ is cyclic. G has finite order if and only if T is elliptic with a rational rotation angle $2\pi p/q$ for $p/q \in \mathbb{Q}$ (where we assume p, q are relatively prime). In this case $G =< S >$, where S is rotation by $2\pi/q$. If the angle of rotation is not of this form, then G is neither discrete nor does it act properly discontinuously on $\mathbb{H}$.*

Proof. The first statement is clear. To prove the second statement, we just have to show that $S \in G$. This is an easy exercise using the Euclidean algorithm. The final statement follows since the orbit of any point other than the fixed point is infinite and lies on a hyperbolic circle which is compact, and so must have accumulation points on the circle. $\square$

In fact a famous result of Weyl from dynamics shows that in the above example, every orbit of an irrational rotation on a circle is dense in the circle.

Proposition 4.12. *(H. Weyl, 1916) Let $T : [0,1] \longrightarrow [0,1]$ defined by mapping $x \longrightarrow x + \alpha$ mod 1. If $\alpha \notin \mathbb{Q}$ then every orbit is dense.*

Proof. First we claim that it is enough to see that there exists a sequence $n_r \in \mathbb{Z}$ such that $T^{n_r}(0) \longrightarrow 0$. In fact if

$$T^{n_r}(0) = n_r\alpha + m_r, \quad m_r \in \mathbb{Z},$$

then

$$T^{n_r}(x) = x + n_r\alpha + m_r.$$

So

$$T^{n_r}(0) - m_r \longrightarrow 0 \Rightarrow T^{n_r}(x) - x = n_r\alpha \text{ mod } 1.$$

So

$$T^{n_r}(x) \longrightarrow x.$$

Now suppose there is no sequence $T^{n_r}(0) \to 0$. Then there exists an open interval $\mathcal{U} \ni 0$ such that

$$T^n(0) \cap \mathcal{U} = \varnothing \quad \forall\, n \neq 0.$$

We claim that, in this case,

$$T^n\left(\frac{\mathcal{U}}{2}\right) \cap T^m\left(\frac{\mathcal{U}}{2}\right) = \varnothing \quad \forall\, n \neq m,$$

(where if $\mathcal{U} = [a,b]$, then $\frac{\mathcal{U}}{2} := \left[\frac{a}{2}, \frac{b}{2}\right]$). If not, there exists

$$x = u + n\alpha \text{ mod } 1 = v + m\alpha \text{ mod } 1, \quad u, v \in \frac{\mathcal{U}}{2}.$$

Then

$$u - v = (m-n)\alpha \text{ mod } 1 \ \in \mathcal{U} \quad \Rightarrow \quad T^{m-n}(0) \in \mathcal{U}.$$

Thus the translates of $\frac{\mathcal{U}}{2}$ are disjoint, but each has positive measure, which is impossible. $\square$

Exercise 4.13. *Let* $T \in \operatorname{Isom}^+ \mathbb{H}$ *be parabolic. If* $S \in \operatorname{Isom} \mathbb{H}$ *commutes with* T*, what can we say about* $< S, T >$*?*

First note that if $T(v) = v$, then $TS(v) = ST(v) = S(v) \Rightarrow S(v) = v$ since T has a unique fixed point. So it easy to see that S is also parabolic with the same fixed point.

Now put the fixed point at ∞. Then $T(z) = z + a$ and $S(z) = z + b$. We have 2 cases depending on whether or not $a/b \in \mathbb{Q}$.
Case (a): $a/b \in \mathbb{Q}$.
We have

$$a/b = n/m, \quad n, m \in \mathbb{Z},\ \operatorname{hcf}(n, m) = 1.$$

So $a = nd$ and $b = md$ with $d \in \mathbb{R}$ and there exist $p, q \in \mathbb{Z}$ such that $pn + qm = 1$.
Let $L : z \mapsto z + d$. Then

$$T = L^n, \quad S = L^m, \quad T^p S^q = L^{np} L^{mq} = L,$$

so $< L >=< S, T >$, that is $< S, T >$ is a *cyclic* group generated by a single parabolic element L.

Case (b): $\alpha = a/b \notin \mathbb{Q}$.
In this case, we claim that every $< S, T >$ orbit in $\mathbb{H}$ has accumulation points in $\mathbb{H}$, so that $< S, T >$ does not act properly discontinuously on $\mathbb{H}$. Pick $z \in \mathbb{H}$. The orbit of z is the set of points

$$T^{n'} S^{m'}(z) = z + n'a + m'b, \quad n', m' \in \mathbb{Z}.$$

We have:

$$\{n'a + m'b\} \quad \text{is dense in } \mathbb{R} \quad \Longleftrightarrow \quad \{n'\alpha + m' : n', m' \in \mathbb{Z}\} \quad \text{is dense in } \mathbb{R}$$

$$\Longleftrightarrow \{n'\alpha \mod 1 :\ n' \in \mathbb{Z}\} \quad \text{is dense in } [0, 1].$$

The last condition is exactly Weyl's theorem above.

It is now easy to see that $< S, T >$ is neither discrete nor acts properly discontinuously. Why?

Lemma 4.14. *Any abelian subgroup* $G \subset \operatorname{Aut} \mathbb{H}$ *which acts properly discontinuously and which contains a parabolic or an elliptic transformation must be cyclic.*

Proof. Let us take the case in which G contains a parabolic element; the other case is similar. As in Exercise 4.13, all elements in G must be parabolic with a common fixed point v. Conjugating to put v at ∞, we may assume all elements are of the form $S_a : z \mapsto z + a, a \in \mathbb{R}$.

Let $A = \{a > 0 : S_a \in G\}$. We claim that $\inf A > 0$. If not, we can find $S_{a_n} \in G$ with $a_n \longrightarrow 0$. It is easy to see that in this case G does not act properly discontinuously since if we take K to be the line interval joining $-1 + i$ to $1 + i$ then $K \cap S_{a_n} K \neq \emptyset$ for infinitely many n. Alternatively this contradicts the discreteness of G, since $\begin{pmatrix} 1 & a_n \\ 0 & 1 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. So by Corollary 4.9, G does not act properly discontinuously.

Now let $\alpha = \inf A$. We claim that $S_\alpha \in G$. Once again, if not, we can find $S_{b_n} \in G$ with $b_n \longrightarrow \alpha$. Then $S_{b_n - b_{n+1}} \in G$ and $b_n - b_{n+1} \longrightarrow 0$ which we have already ruled out.

Now let $T \in G$, so $T = S_a$ or $T^{-1} = S_a$ for some $a \in A$. We can write $a = m\alpha + b$ where $b \in [0, \alpha)$. If $b > 0$ then $TS_\alpha^{-m} = S_b$ so $b \in A$, but if $b > 0$ this contradicts the definition of α. $\square$

The hyperbolic case is slightly more subtle. Suppose that T is hyperbolic and that $ST = TS$. Then S fixes the axis of T, since $S(\mathrm{Ax}\, T) = \mathrm{Ax}\, STS^{-1} = \mathrm{Ax}\, T$.Thus S maps the fixed point set of T to itself. In other words, $S\{v_1, v_2\} = \{v_1, v_2\}$ where v_i are the fixed points of T. However we cannot conclude that $S(v_i) = v_i$. In fact we have:

Lemma 4.15. *Suppose that T is hyperbolic and that $S(\mathrm{Ax}\, T) = \mathrm{Ax}\, T$, so that $S\{v_1, v_2\} = \{v_1, v_2\}$ where v_i are the fixed points of T. Then either:*
(a) $S(v_i) = v_i$, and S is hyperbolic with the same axis, $ST = TS$, and $< S, T >$ is cyclic or
(b) $S(v_i) = v_j$ and S is elliptic of order 2 with fixed point on $\mathrm{Ax}\, T$, $STS^{-1} = T^{-1}$ and $< S, T >$ is a (non-abelian) dihedral group.

Proof. Case (a) follows exactly as in the parabolic and elliptic cases discussed above.

In case (b), we show that the restriction map $S : \mathrm{Ax}\, T \longrightarrow \mathrm{Ax}\, T$ must have a fixed point. There are two methods:
• We can arrange $\mathrm{Ax}\, T$ to be the diameter $[-1, 1]$ of $\mathbb{D}$. Then S restricts to a map of the axis $[-1, 1]$ to itself with $S(1) = -1,\ S(-1) = 1$. So using the intermediate value theorem there must exist a point $x \in \mathrm{Ax}\, T$ such that

$S(x) = x$. So $S^2 = id$ and it follows that S is elliptic of order 2.
• We may as well assume that the fixed points of T are $v_i = \{0, \infty\}$ and that $T(z) = \lambda z$. Let $S = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. Then

$$S(0) = \infty,\ S(\infty) = 0,\ \det S = 1 \ \Rightarrow\ d = 0,\ a = 0,\ c = -1/b,$$

so $S = \begin{pmatrix} 0 & b \\ -\frac{1}{b} & 0 \end{pmatrix}$ which is elliptic of order 2 and has fixed point ib.

In either case, we now note that STS^{-1} is hyperbolic (because T is) and has the same translation length and same axis. However S has interchanged the fixed points and hence the direction of translation. So $STS^{-1} = T^{-1}$ as claimed. □

Putting together the above discussion, we have proved:

Proposition 4.16. *Any discrete abelian subgroup of* $\mathrm{Isom}^+ \mathbb{H}$ *is cyclic.*

Proof. Suppose that $G \subset \mathrm{Isom}^+ \mathbb{H}$ is abelian. We have seen from the examples above that all the elements in G are either all elliptics with the same fixed point; parabolics with a common fixed point; or hyperbolics with a common axis. The arguments above also show that if G is either cyclic or non-discrete. □

The condition of being discrete imposes further restrictions as well. For example:

Lemma 4.17. *Suppose that* $G \subset \mathrm{Isom}^+ \mathbb{H}$ *is discrete, and suppose that* $T \in G$ *is hyperbolic and that* $S \in G$ *is parabolic. Then* S *and* T *cannot have a common fixed point.*

Proof. Without loss of generality we may assume the common fixed point is ∞ so that $T : z \mapsto \lambda z$ and $S : z \mapsto z + c$. Replacing T by T^{-1} if necessary, we may also assume that $|\lambda| > 1$.
Computing we see that:

$$T^{-n}ST^n(z) = \lambda^{-n}(\lambda^n z + c) = z + \lambda^{-n}c.$$

Since $|\lambda| > 1$, we have $T^{-n}ST^n(z) \longrightarrow z$ so $G =< S, T >$ does not act properly discontinuously. Moreover the subset

$$\left\{T^{-n}ST^n = \begin{pmatrix} 1 & \lambda^{-n}c \\ 0 & 1 \end{pmatrix}\right\} \subset \mathrm{PSL}(2, \mathbb{C})$$

is clearly not discrete. □

We end this section with a couple of more complicated examples.

Example 4.18. *Let* $G = \mathrm{SL}(2, \mathbb{Z}) \subset \mathrm{SL}(2, \mathbb{R})$*. This group is clearly discrete.*

Example 4.19. *Start with a tessellation of* $\mathbb{H}$ *and consider its symmetry group. It is fairly clear that this must act properly discontinuously (because only finitely many symmetries map a compact set to itself). By Corollary 4.9 it is discrete as a subgroup of* $\mathrm{PSL}(2, \mathbb{R})$*.*

4.2 Relation of properly discontinuity and discreteness.

In this section we prove the important theorem mentioned above.

Theorem 4.20. $G \in \mathrm{PSL}(2, \mathbb{R})$ *is discrete if and only if its action on* $\mathbb{H}$ *is properly discontinuous.*

Definition 4.21. *A* ***Fuchsian group*** *is a discrete subgroup of* $\mathrm{SL}(2, \mathbb{R})$*.*

We have already proved in Corollary 4.9 that if G acts properly discontinuously, then G is discrete. To prove the converse, we start with the following Lemma.

Lemma 4.22. *Let* $z_0 \in \mathbb{H}$ *and suppose that* $K \subset \mathbb{H}$ *is compact. Then the set* $E = \{T \in \mathrm{PSL}(2, \mathbb{R}) : T(z_0) \in K\}$ *is compact.*

Remark 4.23. *We will only need to use this Lemma in the case* $z_0 = i \in \mathbb{H}$*, in which case the final step of the proof is slightly simpler.*

Proof. Note that a set in $K \subset \mathbb{H}$ is compact if and only if it is closed and bounded. Why? Also note that since $\mathrm{PSL}(2,\mathbb{R})$ inherits the topology from $\mathbb{R}^4$, to see E is compact we also only need see it is closed and bounded.
• (E is closed): Consider the continuous map $\beta : \mathrm{PSL}(2,\mathbb{R}) \longrightarrow \mathbb{H}$ defined by $T \mapsto T(z_0)$. Then $E = \beta^{-1}(K)$ which is closed.
• (E is bounded): Since K is compact, there exists M_1 such that $\forall\ T = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in E$, we have

$$|T(z_0)| = \left|\frac{az_0+b}{cz_0+d}\right| < M_1.$$

Also, as K is compact in $\mathbb{H}$, there exists $M_2 > 0$ such that

$$\Im\left(\frac{az_0+b}{cz_0+d}\right) \geqslant M_2.$$

Using the formula

$$\Im\left(\frac{az+b}{cz+d}\right) = \frac{(ad-bc)\Im z}{(cz+d)^2}$$

and remembering that $ad - bc = 1$ because $T \in \mathrm{PSL}(2,\mathbb{R})$, we deduce that $\Im z_0 > M_2|cz_0+d|^2$ and hence $|az_0+b| < M_1|cz_0+d|$.
So

$$|cz_0+d| < \sqrt{\frac{\Im z_0}{M_2}}, \quad |az_0+b| < M_1\sqrt{\frac{\Im z_0}{M_2}}.$$

This gives upper bounds on $|cz_0+d|$ and $|az_0+b|$. In the case $z_0 = i$ we see easily that $|ci+d|$ and $|ai+b|$ bounded $\Rightarrow |a|, |b|, |c|$ and $|d|$ bounded.

In the general case the same result follows because $|c| > L \Rightarrow |cz_0+d| > |L||z_0| - |d| > k$ as $|L| \longrightarrow \infty$. (Check this.) □

Proof of Theorem 4.20. It remains to show that G discrete implies that G acts on $\mathbb{H}$ properly discontinuously. In other words, given a compact subset $K \subset \mathbb{H}$, we have to show that $T(K) \cap K = \varnothing$ for all but finitely many $T \in G$. Without loss of generality, we can assume that $K = \overline{B_R(i)}$, that is the closed hyperbolic disk with centre i and radius $R >> 0$ (because $K \subset B_R(i)$). (This is because if $T(K) \cap K \neq \varnothing$ then certainly $T(\overline{B_R(i)}) \cap \overline{B_R(i)} \neq \varnothing$.) Also note that $\overline{B_R(i)} \cap T(\overline{B_R(i)}) \neq \varnothing \Rightarrow T(i) \in \overline{B_{2R}(i)}$. In fact, if

$$x = T(y),\ d(x,i) \leqslant R,\ d(y,i) \leqslant R,$$

then

$$d(i, T(i)) \leqslant d(i, T(y)) + d(T(y), T(i)) = d(i, x) + d(y, i) \leqslant 2R.$$

Now let $E_G = \{T \in \mathrm{PSL}(2, \mathbb{R}) : T(i) \in \overline{B_{2R}(i)}\} \cap G$. By Lemma 4.22, the set $\{T : T(i) \in \overline{B_{2R}(i)}\}$ is compact. But G is discrete and so also E_G is discrete. However any infinite subset of a compact set has accumulation points, so E_G must be finite. Thus $T(\overline{B_R(i)}) \cap \overline{B_R(i)} \neq \varnothing$ for all but finitely many $T \in G$ which is what we needed to show. □

Corollary 4.24. *(i) If $G \subset \mathrm{PSL}(2, \mathbb{R})$ is not discrete, then all but possibly one G-orbit have accumulation points.*
(ii) If G is discrete, then no G-orbit has accumulation points in $\mathbb{H}$.
(iii) If G is discrete and infinite, then every orbit of $z_0 \in \mathbb{H}$ has accumulation points on $\partial\mathbb{H}$.

Proof. (i) and (ii) follow immediately from the theorem and Lemma 4.8. (iii) is clear since $\mathbb{H} \cup \partial\mathbb{H}$ is compact and G cannot consist entirely of elliptics with a common fixed point. □

The set of accumulation points of a G-orbit $G(z)$ on $\partial\mathbb{H}$ turns out to be independent of $z \in \mathbb{H} \cup \partial\mathbb{H}$. It is called the *limit set* of G. The limit set is an important object which will be studied in more detail in Chapter 8.

Definition 4.25. *G acts **freely** on X if for any $z \in X, g \in G$ such that $gz = z$, then $g = id$.*

Remark 4.26. *Clearly $G \subset \mathrm{PSL}(2, \mathbb{R})$ acts freely on $\mathbb{H} \iff G$ does not contain elliptic elements.*

Corollary 4.27. *If a discrete group G acts freely, then $\forall\, z \in \mathbb{H}$, there exists a neighbourhood $\mathcal{U}$ of z such that $g\,\mathcal{U} \cap \mathcal{U} = \varnothing\ \forall\, g \in G$.*

Proof. By Corollary 4.24 (ii), z is not an accumulation point of Gz. So let

$$\epsilon = \inf_{g \in G - \{id\}} d(z, gz).$$

Then

$$B_{\frac{\epsilon}{2}}(z) \cap hB_{\frac{\epsilon}{2}}(z) = \varnothing\ \forall\, h \in G, h \neq id.$$

In fact,

$$d(z, hz) \leqslant d(z, w) + d(w, hz) < \epsilon\ \forall\, w \in B_{\frac{\epsilon}{2}}(z) \cap hB_{\frac{\epsilon}{2}}(z).$$

□

Why do we care about properly discontinuous actions? Firstly, because they are the actions associated to tessellation of $\mathbb{H}$. A more high level mathematical reason relates to the theory of covering spaces, which we shall see more of in the final chapter. Recall the following definition:

Definition 4.28. *Given two path connected topological spaces X and Y, a continuous map $p : X \longrightarrow Y$ is a* ***covering map*** *if each $y \in Y$ has a neighbourhood $\mathcal{V}_y$ such that every connected component of $p^{-1}(\mathcal{V}_y)$ is mapped homeomorphically to $\mathcal{V}_y$ by the restriction of the map p.*

Example 4.29. *(i) The map $p_1 : \mathbb{R} \longrightarrow \mathbb{S}^1$ defined by $\theta \mapsto e^{2\pi i\theta}$ is a covering map.*
(ii) The map $\mathbb{R}^2 \longrightarrow \mathbb{T}^2$ defined by $(x, y) \mapsto (e^{2\pi ix}, e^{2\pi iy})$ is a covering map.

Theorem 4.30. *Let G be a group acting on a topological space X. Then X/G is Hausdorff and $p : X \longrightarrow X/G$ is a covering map $\iff$ G acts freely and properly discontinuously on X.*

Proof. See [15], Proposition 3.5.8 p.99. □

4.2.1 More properties of discrete groups

Let us look at some more interesting and useful properties of subgroups $G \subset \mathrm{PSL}(2, \mathbb{R})$.

Lemma 4.31. *Suppose that G is discrete and that $h \in G$ hyperbolic. Then for any $g \in G$, either g has no fixed points in common with h, or g is hyperbolic and has both fixed points in common with h.*

Proof. Without loss of generality, we may suppose that g and h have a common fixed point at $\infty \in \mathbb{H}$ and that $h(z) = \lambda z$ with $|\lambda| > 1$. We also have $g(z) = az + b$ (where $a = 1$ iff g is parabolic). We compute:

$$h^{-n}gh^n(z) = \lambda^{-n}(a\lambda^n z + b) = az + \lambda^{-n}b.$$

This is impossible unless $b = 0$ because G is discrete. □

Lemma 4.32. *Suppose that $g \in Isom^+\mathbb{H}$ is either parabolic or elliptic and $h \in Isom^+\mathbb{H}$ doesn't fix the fixed point of g. Then $ghg^{-1}h^{-1}$ is hyperbolic.*

Proof. • *Method 1*:
By geometry of two reflections, as in Corollary 3.21 in the previous chapter.
• *Method 2*:
Say g parabolic with $g(\infty) = \infty$. Then $g = \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}$ and $h = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ with $c \neq 0$ (so that $h(\infty) \neq \infty$). Hence $ghg^{-1}h^{-1} = \begin{pmatrix} a+ct & b+td \\ c & d \end{pmatrix} \begin{pmatrix} d+tc & -b-ta \\ -c & a \end{pmatrix}$ and so

$$\mathrm{Tr}(ghg^{-1}h^{-1}) = 2ad + c^2t^2 - 2b = 2 + c^2t^2 > 2.$$

The proof of the elliptic case is similar. We may as well assume that the fixed point of g is $0 \in \mathbb{D}$. Then $g = \begin{pmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{pmatrix}$ and $h = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ with $h(0) \neq 0$. Then $b \neq 0$ and $ghg^{-1}h^{-1} = \begin{pmatrix} e^{i\theta}a & e^{i\theta}b \\ e^{-i\theta}c & e^{-i\theta}d \end{pmatrix} \begin{pmatrix} de^{-i\theta} & -be^{-i\theta} \\ -ce^{i\theta} & ae^{i\theta} \end{pmatrix}$.
So

$$\mathrm{Tr}(ghg^{-1}h^{-1}) = ad - bce^{2i\theta} - bce^{-2i\theta} + ad = 2(1+bc) - 2bc\cos 2\theta = 2 + 4bc\sin^2(\theta).$$

Note that $c = \bar{b}, d = \bar{a}$, so

$$\mathrm{Tr}(ghg^{-1}h^{-1}) = 2 + 4|b|^2\sin^2(\theta) > 2.$$

□

Corollary 4.33. *Suppose that G is discrete and that $h \in G$ is hyperbolic. Suppose that there exists $g \in G$ such that $g\,(\mathrm{Ax}\,h) \neq \mathrm{Ax}\,h$. Then G contains infinitely many hyperbolic elements with distinct axes.*

Proof. By the previous two Lemmas, both endpoints of $g\,(\mathrm{Ax}\,h) = \mathrm{Ax}(ghg^{-1})$ are distinct from those of $\mathrm{Ax}\,h$. (We are using the important fact that $T(\mathrm{Ax}\,S) = \mathrm{Ax}\,TST^{-1}$. Why is this true?)
We claim that the set $\{h^n g\,(\mathrm{Ax}\,h)\}$ are an infinite set of axes no two of which have a common end point. In fact, if $h^n g\,(\mathrm{Ax}\,h) = h^m g\,(\mathrm{Ax}\,h)$ with $n \neq m$, then $h^r g\,(\mathrm{Ax}\,h) = g\,(\mathrm{Ax}\,h)$ for some minimum $r > 0$. Let a, b be the endpoints of $\mathrm{Ax}\,h$. Then $h^r g(a), h^r g(b) \in \{g(a), g(b)\}$ and so either $h^r g(a) = g(a), h^r g(b) = g(b)$ or $h^r g(a) = g(b), h^r g(b) = g(a)$. In both case, $h^{2r} g(a) = g(a), h^{2r} g(b) = g(b)$. Thus h^{2r} fixes both $\mathrm{Ax}\,h$ and $g(\mathrm{Ax}\,h)$. But this is impossible, so $r = 0$. □

We have already shown that any abelian Fuchsian group is cyclic. In fact, Fuchsian groups are either extremely simple (virtually abelian) or they contain a free group on at least two generators. The simple examples are known as *elementary* groups

Definition 4.34. *A Fuchsian group G is called **elementary** if it is one of the following list:*
(a) $< T >$ where T is elliptic;
(b) $< T >$ where T is parabolic;
(c) $< T >$ where T is hyperbolic;
(d) the (non-abelian) dihedral group generated by a hyperbolic and an elliptic as in Exercise 4.15 (b).

Note that in each of the first three cases, G fixes either one or two points in $\mathbb{H} \cup \partial\mathbb{H}$. In the last case, G does not fix any points but it does fix a line, namely the axis of T.

Theorem 4.35. *Let G be a Fuchsian group. Then the following are equivalent:*
(i) G is elementary.
(ii) G is virtually abelian, that is there exists a finite index abelian subgroup $H \subset G$.
(iii) G has a finite G-orbit in $\mathbb{H} \cup \partial\mathbb{H}$.

Remark 4.36. *Condition (i) is the easiest to understand; condition (ii) is the most modern and sophisticated and (iii) is sometimes preferred because it is works also in $\mathbb{H}^3$.*

Proof. Proof of (i) $\Rightarrow$ (ii): Obvious.
Proof of (ii) $\Rightarrow$ (iii): Let G be virtually abelian and let H be an abelian subgroup of G so that G/H is finite. Since H is discrete and abelian, we know from Proposition 4.16 that it is cyclic and that there exists v such that $H(v) = v$, that is $H \subset \mathrm{Stab}_G(v)$. Then $|\mathrm{orb}_G(v)| \leq |G/H| < \infty$.
Proof of (iii) $\Rightarrow$ (i):
Suppose first that G contains no hyperbolic elements. Then by Lemma 4.32 all the elements have a common fixed point. We have already analysed these cases and shown that G must be cyclic (either elliptic or parabolic).

Now suppose that G contains a hyperbolic element h. Then either every other element fixes $\mathrm{Ax}\, h$ or not. In the first case either by Proposition 4.16 G is cyclic hyperbolic, or there is an elliptic S of order two with fixed point on $\operatorname{Ax} h$, such that $ShS^{-1} = h^{-1}$. We claim that in this case we are in the dihedral example above. Let H be the set of all hyperbolic elements in G with the same axis as h. By Proposition 4.16 again, H is cyclic, generated by a single hyperbolic h_0 say. Suppose that $H \neq G$. We claim that $[G : H] = 2$. Let $S' \in G, S' \notin H$. Then since S and S' both interchange the endpoints of $\operatorname{Ax} h = \operatorname{Ax} h_0$, $SS' \in G$ fixes the two endpoints of $\operatorname{Ax} h$ and so $SS' \in H$. Thus there are only 2 cosets of H in G, namely H and SH. This proves the claim.

In the second case, there exists $g \in G$ such that $g\,(\operatorname{Ax} h)$ is distinct from $\operatorname{Ax} h$. Moreover by Lemma 4.31 these two axes have no endpoint in common. To finish the proof it will be enough to see that in this case there doesn't exist any finite G-orbit. Clearly, any point in $\mathbb{H} \cup \partial\mathbb{H}$ distinct from the fixed points $h^{\pm}$ of h has an infinite G-orbit. Since $g\,(\operatorname{Ax} h) = \operatorname{Ax} ghg^{-1}$ has no common end with $\operatorname{Ax} h$, the images of h^{+} and h^{-} under powers of ghg^{-1} are all distinct, so the orbits of $h^{\pm}$ are infinite also. This completes the proof. □

4.2.2 General tests for discreteness

Here are some more sophisticated tests for discreteness.

Lemma 4.37. *(Shimuzu's Lemma)*
If G is Fuchsian and $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \in G$, then for any $B = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in G$ we have either $c = 0$ or $|c| \geqslant 1$.

Proof. Let $B_0 = B$ and $B_{n+1} = B_n A B_n^{-1}$. Say $B_n = \begin{pmatrix} a_n & b_n \\ c_n & d_n \end{pmatrix}$. A computation shows that $\begin{pmatrix} a_{n+1} & b_{n+1} \\ c_{n+1} & d_{n+1} \end{pmatrix} = \begin{pmatrix} 1 - a_n c_n & a_n^2 \\ -c_n^2 & 1 + a_n c_n \end{pmatrix}$.
From this $c_{n+1} = -c_n^2$ so that if $|c| < 1$, then $c_n^2 \longrightarrow 0$. Also by induction, we see that $|a_n| \leqslant n + |a_0|$, so $a_n c_n \longrightarrow 0$ and $a_n \longrightarrow 1$. So $B_{n+1} \longrightarrow A$ contradicting discreteness. □

Corollary 4.38. *Under the hypotheses of the previous Lemma, we have that*

$$|(\operatorname{Tr}(ABA^{-1}B^{-1}) - 2| \geqslant 1.$$

Proof. Check that $\operatorname{Tr}[A, B] = 2 + c^2$. □

This is a special case of *Jørgensen's inequality*, the most important result of this type. It applies to subgroups of $\mathrm{SL}(2,\mathbb{C})$.

Theorem 4.39. *(Jørgensen's inequality)*
For any elements A, B in a non-elementary discrete subgroup of $\mathrm{SL}(2,\mathbb{C})$*:*

$$|(\mathrm{Tr}^2(A) - 4| + |(\mathrm{Tr}(ABA^{-1}B^{-1}) - 2| \geqslant 1.$$

Proof. If A is parabolic, this is just the previous corollary. If A is hyperbolic or elliptic, by normalizing we can write $A = \begin{pmatrix} u & 0 \\ 0 & \frac{1}{u} \end{pmatrix}$. Then

$$|(\mathrm{Tr}^2(A) - 4| + |(\mathrm{Tr}(ABA^{-1}B^{-1}) - 2| = (1 + |bc|)\left|u - \frac{1}{u}\right|^2 = \mu.$$

Computing B_n as above, you can show that the sequence has a limit if $|\mu| < 1$. The details are quite complicated, see [10] p. 107. $\square$

Chapter 5

Fundamental domains

5.1 Fundamental domains

Definition 5.1. *If G is a Fuchsian group, then a subset $R \subset \mathbb{H}$ is a* ***fundamental domain*** *for G if:*
(i) if $x, y \in R$ such that $x = gy$ for an element $g \in G$, then $g = id$, equivalently $gR \cap R = \varnothing \; \forall g \in G - \{id\}$.
(ii) $\forall\, z \in \mathbb{H}$, $\exists\, g \in G$ such that $gz \in \overline{R}$, that is

$$\bigcup_{g \in G} \overline{gR} = \mathbb{H}.$$

A similar definition can be for more general groups G acting properly discontinuously on a metric space X. Say R is a fundamental domain for a group G and choose $A \subset R$. It is easy to see that for any $g \neq \mathrm{id}$, the set $(R \setminus A) \cup g(A)$ is also a fundamental domain. Thus a fundamental domain for G is far from being unique.

Definition 5.2. *A subset $C \subset \mathbb{H}$ is called* ***convex*** *if for all x and y in C, every point on the geodesic line segment connecting x and y is in C.*

We shall also usually want to assume the following additional good properties:
(iii) R is the interior of a convex geodesic polygon in $\mathbb{H}$, ie R is the interior of a convex set and $\partial R \cap \mathbb{H}$ is a countable union of geodesic segments of positive length, only finitely many of which meet any compact set.
(iv) $\bigcup_{g \in G} gR$ is ***locally finite****, that is for any compact subset $K \subset \mathbb{H}$, then*

$\sharp\{g : gR \cap K \neq \varnothing\} < \infty.$

Suppose that R is a geodesic polygon. Notice that $\overline{R}$ may not be compact, so that $\partial\mathbb{R}$ may intersect $\partial\mathbb{H}$, in which case the connected components of $\partial\mathbb{R} \cap \partial\mathbb{H}$ may be either isolated points or arcs in $\partial\mathbb{H}$. We call a maximal geodesic segment in ∂R, or a maximal arc in $\partial R \cap \partial\mathbb{H}$, an **edge**. In most cases that we will consider, ∂R will have a finite number of edges.

5.2 Existence of a fundamental domain: Dirichlet and Ford fundamental domains

Theorem 5.3. *Say G is a Fuchsian group acting on $\mathbb{H}$ (or $\mathbb{D}$). Then there exist fundamental domains R for G satisfying (i)-(iv) above. The fundamental domain can be chosen to be finite sided if and only if G is finitely generated.*

We shall prove that such fundamental domains exist by direct construction. There are two main methods: **Dirichlet domains** and **Ford domains**.

We shall show later that if R can be chosen to be finite sided, then G is finitely generated. The converse is somewhat harder and won't be done here, see [10]. It is false in dimension greater than two.

5.2.1 Dirichlet domains

Pick $z_0 \in \mathbb{H}$. For each orbit point $gz_0, g \neq \mathrm{id}$, draw the perpendicular bisector of the line from z_0 to gz_0. This separates $\mathbb{H}$ into two parts. For each $g \in G - \{\mathrm{id}\}$, let $H_g = H_g(z_0) \subset \mathbb{H}$ be the half-space cut out by the perpendicular bisector of z_0 and gz_0 and containing z_0, in other words $H_g(z_0) = \{z \in \mathbb{H} : d(z, z_0) < d(z, gz_0)\}$.

Definition 5.4. *Pick $z_0 \in \mathbb{H}$ such that $gz_0 = z_0 \Rightarrow g = id$. (Why can we do this?). We define the **Dirichlet domain** with centre z_0 to be*

$$R = R_{z_0} = \bigcap_{g \in G - \{id\}} H_g(z_0) = \{z \in \mathbb{H} : d(z, z_0) < d(z, gz_0)\ \forall\ g \neq id\}.$$

We will show that R_{z_0} has the properties (i) – (iv) required by Theorem 5.3. In practice, the half-planes corresponding to a finite number of orbit points "near" z_0 determine R and the remaining half spaces H_{g_n} don't cut down R further. The construction depends heavily on z_0: we shall see some actual examples below.

Proof that R_{z_0} satisfies the conditions of Theorem 5.3

♦ We begin by showing that $h(R_{z_0}) = R_{hz_0}$:

$$\begin{aligned} h(R_{z_0}) &= \{hz \in \mathbb{H} : d(z, z_0) < d(z, gz_0)\ \forall\ g \neq id\} \\ &= \{hz \in \mathbb{H} : d(hz, hz_0) < d(hz, hgz_0)\ \forall\ g \neq id\} \\ &= \{w \in \mathbb{H} : d(w, hz_0) < d(w, hgh^{-1}hz_0)\ \forall\ g \neq id\} \\ &= \{w \in \mathbb{H} : d(w, hz_0) < d(w, khz_0)\ \forall\ k \neq id\} = R_{hz_0}, \end{aligned}$$

where for the last equality we used the obvious fact that

$$\{hgh^{-1} \in G : g \neq id\} = \{k \in G : k \neq id\}.$$

♦ *Proof of (i)*:
If $z \in R_{z_0} \cap R_{gz_0}$, then $d(z, z_0) < d(z, gz_0)$ and $d(z, gz_0) < d(z, g^{-1}gz_0) = d(z, z_0)$. So

$$R_{z_0} \cap R_{gz_0} = R_{z_0} \cap g(R_{z_0}) = \varnothing.$$

♦ *Proof of (ii)*:
The cardinality $\sharp\{Gz_0 \cap B_1(z)\}$ is finite by the properly discontinuity of the action of the group G. So there exist only finitely many nearest orbit points $g_1z_0, \ldots, g_nz_0$ to z. So we can choose an element g such that $d(z, gz_0) = \inf_{h \in G} d(z, hz_0)$, that is

$$d(z, gz_0) \leqslant d(z, hz_0)\ \forall\ h \neq g.$$

Hence

$$d(g^{-1}z, z_0) \leqslant d(g^{-1}z, g^{-1}hz_0)\ \forall\ h \neq g.$$

We would like to conclude that this gives $g^{-1}z \in \overline{R}$, that is

$$d(w, z_0) \leqslant d(w, kz_0)\ \forall\ k \neq id \tag{5.1}$$

implies that $w \in \overline{R}$. We will do this by showing that (5.1) implies that the half open line segment $[z_0, w) \subset R$, from which it follows immediately that $[z_0, w] \subset \overline{R}$.

If $\alpha \in [z_0, w)$, then $d(z_0, \alpha) + d(\alpha, w) = d(z_0, w) \leq d(w, hz_0)\ \forall\ h \neq id$, where the last inequality follows from (5.1). Thus $d(z_0, \alpha) \leq d(w, hz_0) - d(\alpha, w) \leq d(\alpha, hz_0)\ \forall\ h \neq id$. Moreover the second inequality is an equality only if α is on the line from hz_0 to w. Since this line intersects $[z_0, w)$ exactly in w, this only happens if $\alpha = w$, which by our assumption is not the case. Hence $d(z_0, \alpha) < d(\alpha, hz_0)\ \forall\ h \neq id$ hence $\alpha \in R$ as claimed.

⧫ *Proof of (iv)*:

Without loss of generality, we can take K to be the closed disk of radius R and centre z_0. Then we have that $gR_{z_0} \cap K \neq \varnothing \Rightarrow \exists z \in gR_{z_0} = R_{gz_0}$ such that $d(z, z_0) \leqslant R$. Let $z = gw$ with $w \in R_{z_0}$. Then

$$d(gz_0, z_0) \leqslant d(z, gz_0) + d(z, z_0) \leqslant d(z, z_0) + d(z, z_0) \leqslant 2R$$

because $z \in R_{gz_0}$. Since there are only finitely many orbit points within distance $2R$ from z_0, the result follows.

⧫ *Proof of (iii)*:

It is clear that R is convex since each half space H_g is convex and the intersection of convex sets is convex.

We claim that if K is a compact set, then only finitely many common perpendiculars meet K. Again without loss of generality we can take K to be the closed disk of radius R and centre z_0. Say $w \in L_h$, the common perpendicular to z_0 and hz_0, and $w \in K$. Then

$$d(z_0, hz_0) \leqslant d(z_0, w) + d(w, hz_0) = 2d(w, z_0) \leqslant 2R.$$

The claim follows from (iv).

Thus if B is any open ball with $R \cap B \neq \emptyset$, then only finitely many common perpendiculars meet B so that $\partial R \cap B$ contains at most finitely many geodesic segments. To complete the proof of (iii), it is enough to show that R is open, equivalently that $R \cap B$ is open for any open ball B.

Assume that $R \cap \overline{B} \neq \varnothing$ (otherwise there is nothing to prove). So $H_g \cap \overline{B} = H_g$ for all but finitely many $g \in G$, because $\overline{B} \subset H_g$ for all but finitely many $g \in G$. So

$$R \cap \overline{B} = [\bigcap_{g \in G} H_{g \in G}] \cap \overline{B} = (H_{g_1} \cap \cdots \cap H_{g_n}) \cap \overline{B}.$$

So R is open in $\overline{B}$, (and it is impossible for the half lines L_g to accumulate on R).

5.2.2 Ford domains

For this construction, we will assume that G contains parabolic elements which fix ∞. We let $G_0 =< T >$ be the subgroup which stabilises ∞, where $T(z) = z + t$. Note that if $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in G \setminus G_0$, then $c \neq 0$. For later use, we remark that it follows easily from Shimuzu's lemma 4.37, that in this situation there is a positive lower bound on $|c|$. If $t = 1$, the lemma says that $|c| > 0$. If $t \neq 1$, let $S(z) = \lambda z$. Check that $STS^{-1}(z) = z + \lambda t$ so taking $\lambda = 1/|t|$, we can apply the lemma to the group SGS^{-1}. Since $SgS^{-1} = \begin{pmatrix} a & \lambda b \\ \lambda^{-1} c & d \end{pmatrix}$ we deduce that $|c| \geq 1/|t|$.

This Ford domain construction uses isometric circles, see Section 3.1.5. Remember that if $g \in \mathrm{PSL}(2, \mathbb{C})$, then $g'(z) = \dfrac{1}{(cz+d)^2}$. If $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{PSL}(2, \mathbb{C})$ with $c \neq 0$, then the **isometric circle** I_g of g is the circle with centre $-\frac{d}{c}$ and radius $\frac{1}{|c|}$:

$$I_g = \{z \in \mathbb{C} : |cz + d| = 1\} = \{z \in \mathbb{C} : |g'(z)| = 1\}.$$

We define the region inside I_g as

$$D_g = \{z \in \mathbb{C} : |cz + d| < 1\}$$

and the region outside I_g as

$$E_g = \{z \in \mathbb{C} : |cz + d| > 1\}.$$

Observe that:
(i) $g|_{D_g}$ is an expansion (because $|g'(z)| > 1$).
(ii) $g|_{E_g}$ is a contraction (because $|g'(z)| < 1$).

Using the chain rule is an easy exercise to prove the following facts:

Exercise 5.5. *(i)* $g(I_g) = I_{g^{-1}}$;
(ii) $g(E_g) = D_{g^{-1}}$
(iii) $g(D_g) = E_{g^{-1}}$.

Definition 5.6. *Let $\Sigma \subset \mathbb{H}$ be any vertical strip $c < \Re z < t + c, c > 0$ of width t. We define the* ***Ford domain*** *as*

$$R = \Sigma \cap \bigcap_{g \in G - \{G_0\}} E_g.$$

Proof that R satisfies the conditions of Theorem 5.3

♦ *Proof of (i)*:
If $z \in R$, then $z \in E_g$ with $g \notin G_0$. So $g(z) \in D_{g^{-1}}$, hence $g(z) \notin R$.
So $z \in gR \cap R \iff z \in R, g^{-1}z \in R$, but $g^{-1}z \in D_g$ and $R \cap D_g = \varnothing$. So $gR \cap R = \varnothing \; \forall g \in G - \{id\}$, as we wanted to prove.
♦ *Proof of (ii)*:
Let $z \in \mathbb{H}$. Clearly by applying suitable element in G_0 we may assume that $z \in \overline{\Sigma}$.

Consider $\Im g(z)$ with $g \in G$. We claim that either $z \in R$ or there exists $g \in G$ such that $\Im g(z) > \Im z$. In fact, if $z \notin R$, then $z \in D_g$ for some $g \in G$, $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ with $c \neq 0$. So $|cz + d| < 1$ and so $\Im g(z) = \frac{\Im z}{|cz+d|^2} > \Im z$.

Now using the above observation on Shimuzu's lemma, for any $g \in G \setminus G_0$ we have $|c| \geq 1/|t|$ so that $\mathrm{rad}(I_g) = \frac{1}{|c|} \leqslant |t|$. Thus if $\Im z > |t|$, then $z \in R$.

In this way, we can inductively find a sequence $g_n \in G \setminus G_0$ such that $\Im g_n(z)$ is strictly increasing and $g_n(z) \in \Sigma$. The sequence either terminates with $g_n(z) \in R$ for some n, or is infinite. But in the second case we would have an accumulation point of the orbit of z in a compact subset of $\mathbb{H}$, which is impossible. □

Exercise 5.7. The Ford domain can be viewed as the limit of the Dirichlet domain as the base point $z_0 \longrightarrow \infty$.

Notice first that if $z_0 \in \mathbb{H}$, then the common perpendicular L_g to the line from z_0 to gz_0 can be described as follows. Let $C_R(z_0)$ be the ball of radius R centre z_0. There is a unique R for which the circles $C_R(z_0)$ and $g(C_R(z_0)) = C_R(gz_0)$ are tangent. Then L_g is the geodesic through the point of tangency and tangent to the two circles.

Now let $z_0 \longrightarrow \infty \in \partial\mathbb{H}$. The circle $C_R(z_0)$ converges to a horocycle $H(d) = \{z \in \mathbb{H} : \Im z = d\}$ for some $d > 0$. There is a unique value of d such that $g^{-1}(H(d))$ (which is a circle tangent to $\mathbb{R}$ at $g^{-1}(\infty)$) is tangent to $H(d)$. There is a unique geodesic through the tangency point in the tangent direction. Prove that this geodesic is the isometric circle of g.

5.3 Examples of fundamental domains

5.3.1 Dirichlet domains for the elementary groups

For the elementary groups $G =< T >$ the Dirichlet domains are as follows:
(i) Elliptic case:
Without loss of generality, we may assume that T is the rotation by $\frac{2\pi}{n}$ about $0 \in \mathbb{D}$. Let $z_0 = \rho_0 e^{i\theta_0}, \rho \neq 0$. The Dirichlet region centre z_0 is the sector

$$\{z = \rho e^{i\theta} : \theta_0 - \pi/n < \theta < \theta_0 + \pi/n\}.$$

(ii) Parabolic case:
Normalize so that T is the parabolic transformation which fixes ∞, that is $T(z) = z + a$ for some $a > 0$. The Dirichlet region centre z_0 is the strip

$$\{z : \Re z_0 - a/2 < \Re z < \Re z_0 + a/2\}.$$

(iii) Hyperbolic case:
Say $T(z) = a^2 z$ for some $a > 1$. The Dirichlet domain centre z_0 is the region between two semicircles:

$$\{z : |a^{-1} z_0| < |z| < |a z_0|\}.$$

5.3.2 Dirichlet domains for the action of $\mathbb{Z}^2$ on $\mathbb{R}^2$

Let $X = \mathbb{R}^2$ and let $G = \mathbb{Z}^2$ act on X by horizontal and vertical translations by 1. Thus if we define $S : (x, y) \mapsto (x + 1, y)$ and $T : (x, y) \mapsto (x, y + 1)$, then $G = \{S^n T^m : n, m \in \mathbb{Z}\}$.

The Dirichlet domain construction still makes sense. The Dirichlet domain centre $z_0 \in X$ is the unit square with z_0 at its centre. Observe that there are *three* common perpendiculars through each vertex of R.

Exercise 5.8. *Experiment with finding Dirichlet domains with different centres for the group $\mathbb{Z}^2$ acting on $\mathbb{R}^2$ as above. In fact, if G acts a lattice subgroup, $S : (x, y) \mapsto (x + 1, y)$ and $T : (x, y) \mapsto (x + a, y + b)$ with $b > 0$, then, in general, the Dirichlet region is a hexagon.*

5.3.3 Dirichlet domains for the action of $\mathrm{SL}(2,\mathbb{Z})$ on $\mathbb{H}$

Let the group $G = \mathrm{SL}(2,\mathbb{Z})$ acting on the space $X = \mathbb{H}$. We will calculate the Dirichlet domain with centre $z_0 = 2i$.
Let $J(z) = -\frac{1}{z}$ and $S(z) = z + 1$, then $J(2i) = \frac{i}{2}$, $S(2i) = 2i + 1$ and $S^{-1}(2i) = 2i - 1$.
(Recall that in the Dirichlet domain construction, we need that if $gz_0 = z_0$, then $g = id$. So, for example, in this case we can't choose $z_0 = i$ as centre because $J(i) = i$, while $z_0 = 2i$ is allowed.)
Calculating we see that $J(2i) = \frac{-1}{2i} = \frac{i}{2}$, $S(2i) = 2i+1$ and $S^{-1}(2i) = 2i-1$.

Proposition 5.9. *The region $R = \{z \in \mathbb{H} : |\Re z| < \frac{1}{2}, |z| > 1\} \subset \mathbb{H}$ is the Dirichlet region with centre $2i$ for the action of $G = \mathrm{SL}(2,\mathbb{Z})$ on $\mathbb{H}$.*

Proof. The perpendicular bisector of the line from $2i$ to $i/2$ is the geodesic $|z| = 1$; likewise the perpendicular bisectors of the lines from $2i$ to $2i \pm 1$ are the geodesics $\Re z = \pm 1/2$. The half planes associated to these three lines cut out the region R. Thus if Δ denotes the Dirichlet domain R_{2i}, then certainly $\Delta \subset R$, so in particular if $z \in \mathbb{H}$ then there exists $g \in G$ with $gz \in R$.

To show that $\Delta = R$, it will therefore suffice to show that $gR \cap R = \varnothing$ $\forall g \in G - \{id\}$. So suppose that $gR \cap R \neq \varnothing$ for some $g \in G - \{id\}$. Then there exists a point $z \in R$ such that $g^{-1}z \in R$ also.
Let $g^{-1} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. Then $\Im g^{-1}(z) = \frac{\Im z}{|cz+d|^2}$.

Note that if $z \in R$, then

$$|cz + d|^2 = c^2|z|^2 + 2|cd|\Re z + d^2 > c^2 - |cd| + d^2 = (|c| - |d|)^2 + |cd| \geqslant 1$$

since $c, d \in \mathbb{Z}$ and they are not both 0. Hence $\Im g^{-1}(z) = \frac{\Im z}{|cz+d|^2} < \Im z$.
Making the same argument with the pair of points $g^{-1}z$ and $gg^{-1}z = z$ and replacing g^{-1} with g, we see equally that $\Im z < \Im g^{-1}z$.
So we obtain $\Im z < \Im g^{-1}z < \Im z$ which is absurd. Thus $gR \cap R = \varnothing$ $\forall g \in G - \{id\}$ as claimed.

□

Of course this method is very special to $\mathrm{SL}(2,\mathbb{Z})$.

Exercise 5.10. *Show that the Dirichlet region for* $\mathrm{SL}(2,\mathbb{Z})$ *with centre* $z_0 = i + \frac{1}{2}$ *is the region* $R' = \{z \in \mathbb{H} : |z| > 1, |z - 1/2| > 1, 0 < \Re z < 1\}$.

Hint: use the element $\Omega = \begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix}$ *which is elliptic of order* 3*. You can see directly that* R' *is a fundamental domain by cutting and pasting the fundamental domain* R *above.*

5.3.4 Ford domains for the action of $\mathrm{SL}(2,\mathbb{Z})$ on $\mathbb{H}$

Let $S = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ and $J = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \in \mathrm{SL}(2,\mathbb{Z})$ as above.

The group contains the element $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ that fixes ∞, so, as we have explained above, we have to restrict the Ford domain $\bigcap_{g \in G - \{id\}} E_g$ to a strip Σ of width 1, that is $R = \Sigma \cap \left(\bigcap_{g \in G - \{id\}} E_g \right)$.

The isometric circle of $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ has centre $-\frac{d}{c}$ and radius $\frac{1}{|c|}$ with $c \neq 0$. Observing that all the entries of the matrix are in $\mathbb{Z}$, then the biggest radius is 1. Hence the radii are $\leqslant 1$ and the equality holds if and only if $|c| = 1$. So if $|c| \neq 1$, then the radius is $\leqslant \frac{1}{2}$; while if $|c| = 1$, then $-\frac{c}{d} \in \mathbb{Z}$.

Consider $J = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$, then I_J has centre 0 and radius 1.

Also the element $N = \begin{pmatrix} 0 & -1 \\ 1 & d \end{pmatrix} \in \mathrm{SL}(2,\mathbb{Z})$ for every $d \in \mathbb{Z}$, so here are isometric circles of radius 1 with centre any integer. Depending on how we choose the strip Σ, we get the same fundamental domains R or R' as in Section 5.3.3.

Chapter 6

Side pairing and the cycle condition

In this chapter we shall answer several questions at the same time: How to create more discrete groups, how to find presentations for these groups, and how to recognize when a given polygon is a fundamental domain for a particular group.

6.1 Side pairings

We begin by examining some basic features of fundamental domains. Suppose that R is a finite sided convex geodesic polygon which is a fundamental domain for some discrete group $\Gamma \subset \mathrm{Isom}^+\mathbb{H}$. Certain copies of R are adjacent to R. Of particular interest are those copies which are adjacent to R along a segment of positive length (as opposed to those copies which meet $\overline{R}$ simply in a vertex).

Definition 6.1. *A* ***side*** *of R is a segment $s \subset \partial R$ of positive length such that $s \subset \overline{R} \cap g(\overline{R})$ for some $g \in \Gamma$.A* ***side pairing*** *of R is an element $g \in \Gamma$ such that $\overline{R} \cap g(\overline{R})$ is a side.*

Notice that sides are not necessarily maximal geodesic segments in ∂R. Maximal segments in ∂R are called **edges** of R). Thus an edge may be a union of several sides. Note also that the definition of side-pairings allows a side to be paired to itself.

In what follows, we will start by assuming that R is a fundamental polygon for some discrete group Γ and $g \in \Gamma$. However later on we shall simply suppose that R is a polygon and we are given a set of isometries which pair its sides. It turns out that, under suitable conditions, these side pairings will generate a discrete group for which R is the fundamental domain. This is part of the famous Poincaré theorem which we will be discussing in this and the next chater.

In these notes, it will be important to associate a side-pairing transformation to each a side of R in a well defined way. *We adopt the convention that the side-pairing transformation associated to a side s of $\overline{R}$ is the unique element $g_s \in \Gamma$ which carries s to another side of $\overline{R}$.* This is illustrated in Figure 6.1. Suppose that $g^{-1}\overline{R}, \overline{R}$ are adjacent along a side s. Then g carries the pair of regions $g^{-1}\overline{R}, \overline{R}$ to the pair $\overline{R}, g\overline{R}$ and hence carries $s = g^{-1}\overline{R} \cap \overline{R}$ to $g(s) = \overline{R} \cap g\overline{R}$. Hence by our convention, the side-pairing associated to s is g, i.e. $g = g_s$.

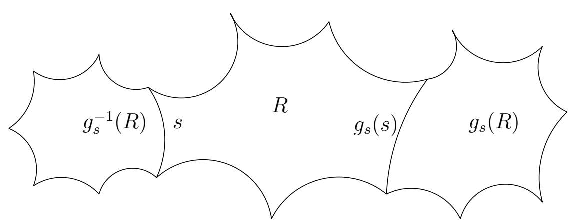

Figure 6.1: Side-pairings: g_s carries $g_s^{-1}R$ to R and s' to s.

The following lemma records the basic facts about side pairings, as illustrated in Figure 6.1.

Lemma 6.2. *Let s be a side of R and with associated side-pairing g_s. Then:*
(i) $s = \overline{R} \cap g_s^{-1}\overline{R}$ and $g_s(s) = \overline{R} \cap g_s\overline{R}$;
(ii) If g_s is a side pairing, then so is g_s^{-1}. Moreover g_s^{-1} is the side-pairing associated to the side $g_s(s)$.

Proof. Referring to Figure 6.1, this follows easily from our definitions and the discussion above. □

We write

$$\Gamma^* = \{g \in \Gamma : g \text{ is a side pairing of } R\}.$$

It follows from the lemma that Γ^* is symmetric, that is, if $g \in \Gamma^*$, then $g^{-1} \in \Gamma^*$.

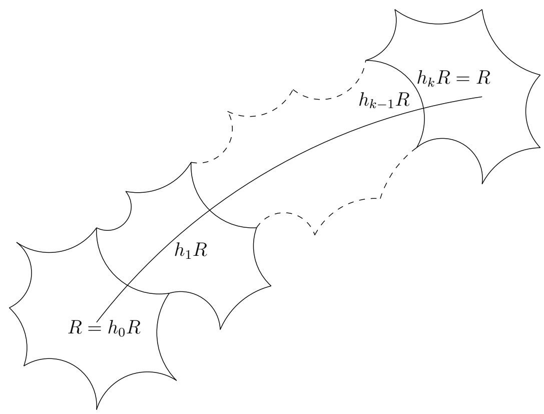

Figure 6.2: Path across the tessellation from $z \in R$ to $hz \in hR$.

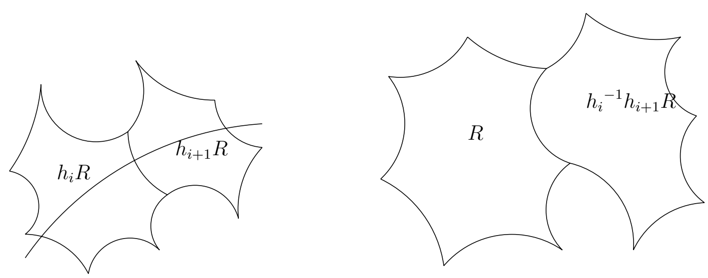

Figure 6.3: Adjacent regions along the path β and the 'pull backs' by h_i^{-1}.

Proposition 6.3. *Suppose that Γ is a Fuchsian group and that R is a finite sided convex geodesic polygon which is a fundamental domain for Γ. Suppose that for each side $s \in R$ there exists $g \in \Gamma$ which pairs s to some side s' of R. Then $\Gamma^* = \{g \in \Gamma : g \text{ is a side pairing of } R\}$ generates Γ, that is for every $h \in \Gamma$ then $h = g_{i_1} \cdot g_{i_2} \cdots g_{i_n}$ where $g_{i_r} \in \Gamma^*$.*

Proof. Let $h \in \Gamma$. Choose $z \in R$ and consider the geodesic path β from z to hz. Adjust β slightly if necessary so that it avoids all the vertices of the tiles $g(R)$, where $g \in \Gamma$, so we get the arrangement in Figure 6.3. Label the copies of R crossed by β in order $R = R_0, R_1, \ldots, R_n = h(R)$. Let $R_i = h_i R$ for $i = 0, \ldots, n$ with $h_i \in \Gamma$, $h_0 = id$ and $h_n = h$. Notice that from the definition, each pair of regions $h_i R, h_{i+1} R, i = 0, \ldots, n-1$ are adjacent along a common side.

If we apply h_i^{-1} to the adjacent regions $h_iR, h_{i+1}R$ we get the regions $R = R, h_i^{-1}h_{i+1}R$. Thus these two regions are also adjacent, so that $h_i^{-1}h_{i+1} \in \Gamma^*$. We have $h = h_n = \prod_{i=0}^{n-1} h_i^{-1}h_{i+1}$ which proves the result. (Observe that, in general, this expression for h is not unique.) □

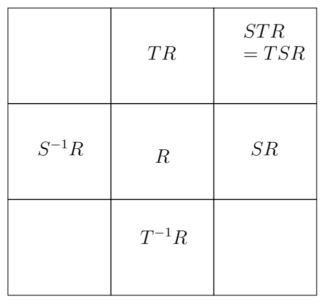

Figure 6.4: Copies of the fundamental domain for $\mathbb{Z}^2 =< S, T >$ meeting round the basic fundamental domain R.

Example 6.4. Consider the space $X = \mathbb{R}^2$ and the group $G = \mathbb{Z}^2$ acting on X by horizontal and vertical translations by 1. Thus G is generated by $S : (x, y) \mapsto (x + 1, y)$ and $T : (x, y) \mapsto (x, y + 1)$, in fact $G = \{S^nT^m : n, m \in \mathbb{Z}\}$. The unit square $R = \{(x, y) \in \mathbb{R}^2 : 0 < x < 1, 0 < y < 1\}$ is a fundamental domain for G. The side pairings are S, T, S^{-1} and T^{-1} and obviously they generate G. The arrangement of SR, TR, $S^{-1}R$ and $T^{-1}R$ round ∂R is shown in Figure 6.4.

6.1.1 Relations between the side pairings: the neighbourhood of a vertex of R

We have learnt that the discrete group Γ associated to a tessellation is generated by the side pairings of a fundamental domain. To understand Γ as an abstract group we need not only a set of generators but also a set of relations between the generators, which between them imply all possible other relations. Such a collection of generators and relations is called a **presentation** of G. It turns out that a set of relations which give a presentation for Γ

can be read off by studying what happens near the vertices of a fundamental domain, as we will explain.

Suppose we are given a finite sided geodesic polygon R which is a fundamental domain for the action of a discrete group Γ, so that the images of R form a locally finite tessellation of $\mathbb{H}$, together with a collection of side pairings which match sides of R in pairs. (Remember that we allow a side to be paired to itself.)

Consider the set of copies of R which meet round one of its vertices v, as shown in Figure 6.5. By local finiteness, only finitely many images of R meet at v. Thus if we label the regions in clockwise order round v, then eventually we come back to where we started. In other words, we can label the regions $h_0R = R, h_1R, \ldots h_{k-1}R, h_kR = h_0R = R$, where $h_0 = h_k = id$.

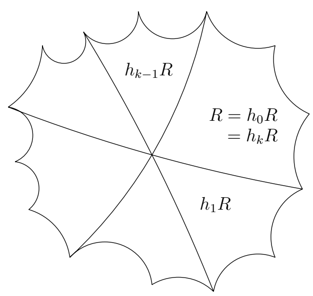

Figure 6.5: Copies of R meeting round a vertex v.

By the same reasoning as in the proof of the Proposition 6.3, we have:
(i) $id = h_k = (h_0^{-1}h_1)\cdots(h_{k-1}^{-1}h_k)$;
(ii) $h_{i-1}^{-1}h_i \in \Gamma^*$ for $i = 1, \ldots, k$.
It will be convenient to rewrite this slightly: if we set $e_i^{-1} := h_i^{-1}h_{i+1} \in \Gamma^*$ for $i = 0, \ldots, k-1$, then we can re-write (i) as $id = h_k = e_0^{-1}\cdots e_{k-1}^{-1}$, or equivalently $id = e_{k-1}\cdots e_0$. Thus we have a relation among the side pairings, which we call a **cycle relation** or a **vertex cycle**. If we start at a different vertex of R, the relation we get may be quite distinct, or it may only differ from the one we already found by cyclic permutation, which one can also think of as conjugation: $id = e_{k-1}\cdots e_0$ is the same relation as $id = e_0e_{k-1}\cdots e_1$ since $e_0e_{k-1}\cdots e_1 = (e_0)e_{k-1}\cdots e_0(e_0^{-1})$. We shall shortly see many examples of how this works.

The following important and much harder result is part of **Poincaré's theorem** which we will prove in Chapter 7.

Proposition 6.5. *Taking these relations for all vertices of R and choosing one relation from each cyclic conjugacy class, gives a presentation for* $\Gamma =< \Gamma^* >$.

Example 6.6. Although the above results are stated for discrete groups acting on $\mathbb{H}$, they apply equally to a discrete group acting on $\mathbb{R}^2$. Thus to understand cycle relations better, let us look at our old example of the space $X = \mathbb{R}^2$ and of the group $\Gamma = \mathbb{Z}^2$ acting on X by horizontal and vertical translations given by the maps $S : (x, y) \mapsto (x + 1, y)$ and $T : (x, y) \mapsto (x, y + 1)$. We proved before that a fundamental domain for Γ is the unit square $R = \{(x, y) \in \mathbb{R}^2 : 0 < x < 1, 0 < y < 1\}$ and that the side pairings are S, T, S^{-1} and T^{-1}, in other words, $\Gamma^* = \{S^{\pm 1}, T^{\pm 1}\}$. Clearly $\Gamma = \{S^n T^m : n, m \in \mathbb{Z}\} =< \Gamma^* >$.

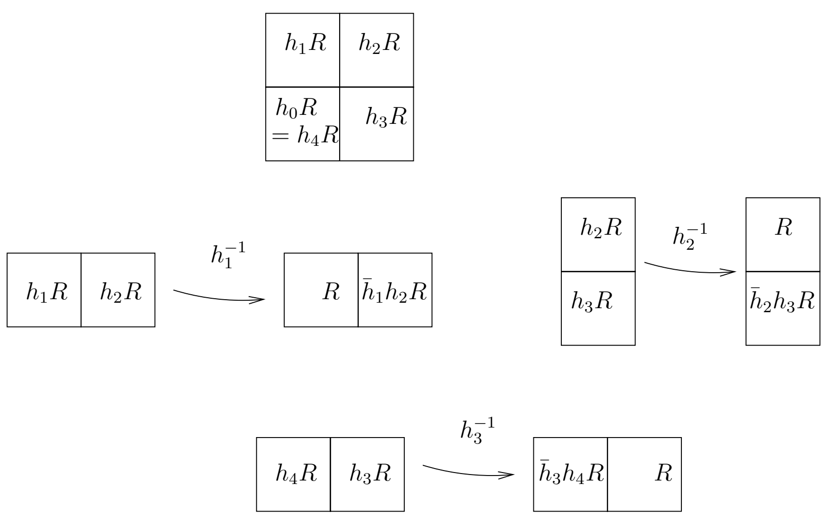

Figure 6.6: Four copies of R round the upper right vertex of R. The lower pictures illustrate how we pull back adjacent pairs of regions so that one of the pair is R. To save space we have written $\bar{h}$ for h^{-1}.

Now we refer to Figure 6.6, in which we look at the copies of R meeting at its top right hand vertex. Going round this vertex starting from the bottom left square R we have:
(i) $h_0 = id$;
(ii) $h_1 = T$;
(iii) $h_2 = TS$. To see this, we apply h_1^{-1} to the adjacent squares h_1R, h_2R. Since the square immediately to the right of R is SR, we deduce $h_1^{-1}h_2 = S$. Thus $h_2 = h_1S = TS$.
(iv) $h_3 = TST^{-1}$. To see this, we apply h_2^{-1} to the adjacent squares h_2R, h_3R. Since the square immediately below R is $T^{-1}R$, we deduce $h_2^{-1}h_3 = T^{-1}$ and so $h_3 = h_2T^{-1} = TST^{-1}$.
(v) $h_4 = id = TST^{-1}S^{-1}$. To see this, we apply h_3^{-1} to the adjacent squares h_4R, h_3R. We see that $h_3^{-1}h_4 = S^{-1}$ and so $h_4 = h_3S^{-1} = TST^{-1}S^{-1}$. Thus we conclude that $id = TST^{-1}S^{-1}$, in other words, $TS = ST$. Clearly in this case, $\Gamma = \mathbb{Z}^2 =< S, T | TS = ST >$, as claimed in Proposition 6.5.

We remark that with our previous notation we have: (i) $e_0 = h_1^{-1}h_0 = T^{-1}$; (ii) $e_1 = h_2^{-1}h_1 = S^{-1}T^{-1}T = S^{-1}$; (iii) $e_2 = h_3^{-1}h_2 = S^{-1}TS = T$; (iv) $e_3 = h_4^{-1}h_3 = S$, so that $e_i \in \Gamma^*$ and $e_0e_1e_2e_3 = id$.

6.1.2 Reading off the relations mechanically.

As before, let R be the fundamental domain for a discrete group G. We want to find a mechanical way of reading off the vertex relations in G. Choose a starting vertex v_0 and let s_0 be the side of R which ends in v_0 and which comes before v_0 in clockwise order round ∂R, as in Figure 6.7.

Let g_0 be the side pairing associated to s_0. Let $v_1 = g_0(v_0)$. Now let s_1 be the side immediately before v_1 in the clockwise order round ∂R. Let g_1 be the side-pairing associated to s_1, and let $v_2 = g_1(v_1)$. Continuing this process, we obtain a sequence:

$$v_0 \xrightarrow{g_0} v_1 \xrightarrow{g_1} v_2 \xrightarrow{g_2} \cdots v_r \xrightarrow{g_r} \cdots$$

Eventually the sequence of vertices will repeat, because there are only finitely many vertices, and because the sequence goes backwards as well as forwards. We shall discuss the point at which we stop the cycle after we have looked at the procedure in more detail below.

To see the connection between this procedure and our previous way of reading off regions round vertices, we proceed as follows. As above, we define

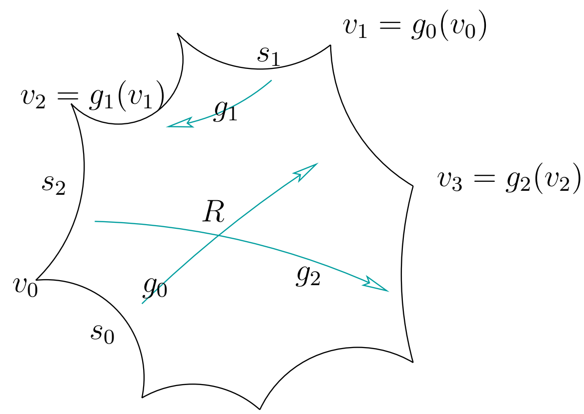

Figure 6.7: Starting the procedure for finding the vertex cycles.

$e_i = h_{i+1}^{-1} h_i$ so that $e_i e_{i-1} \dots e_0 = h_{i+1}^{-1}$.

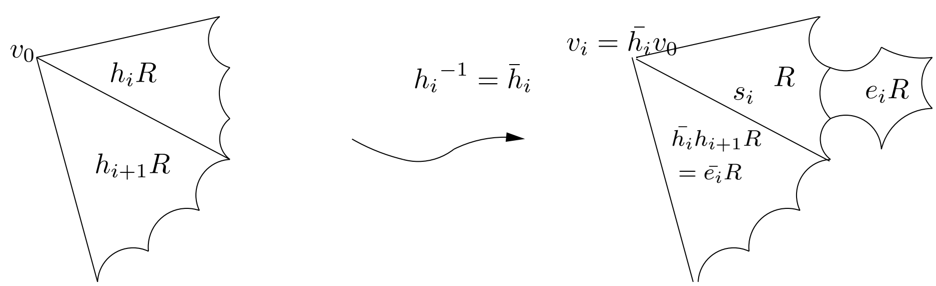

Figure 6.8: Action of $h_i^{-1} = \bar{h}_i$ on adjacent regions meeting at v_0.

We started at v_0. We applied the side pairing g_0 associated to the side s_0 immediately before v_0. The copies of R adjacent along this side are $R = h_0 R$ and $h_1 R$, where $h_0 = id$. Thus $g_0 = h_1^{-1} = e_0$ and by definition $v_1 = e_0 v_0$.

Now we need to know the side immediately before v_1 in the clockwise order round ∂R. We can find this by applying h_1^{-1} to the adjacent pair $h_1 R, h_2 R$ as shown in Figure 6.8 (with $i = 1$). Under this map, the vertex

v_0 maps to $h_1^{-1}v_0 = e_0v_0 = v_1$. Thus the right hand side of the figure (again with $i = 1$) illustrates exactly the configuration we need to read off the side s_1 immediately before v_1 in the clockwise order round ∂R. We see that s_1 is the intersection of the adjacent regions R and $h_1^{-1}h_2R$. Since $h_1^{-1}h_2 = e_1^{-1}$, the relevant side pairing is $g_1 = e_1$. Thus $v_2 = e_1v_1 = e_1e_0v_0 = h_2^{-1}v_0$.

Now suppose inductively that $v_{i-1} = h_{i-1}^{-1}v_0$ and that the side pairing associated to the side before v_{i-1} is e_{i-1}. Then $v_i = e_{i-1}v_{i-1} = e_{i-1}e_{i-2}\dots e_0v_0 = h_i^{-1}v_0$. So we can read off from Figure 6.8 that the side pairing associated to s_i is $h_{i+1}^{-1}h_i = e_i$ and hence that $v_{i+1} = e_iv_i = h_{i+1}^{-1}v_0$.

We conclude that the above sequence of vertices and side-pairings

$$v_0 \xrightarrow{g_0} v_1 \xrightarrow{g_1} v_2 \xrightarrow{g_2} \cdots v_r \xrightarrow{g_r} \cdots$$

can be written

$$v_0 \xrightarrow{e_0} e_0v_0 = v_1 \xrightarrow{e_1} v_2 \xrightarrow{e_2} \cdots v_{k-1} \xrightarrow{e_{k-1}} v_k = v_0$$

where $e_i = h_{i+1}^{-1}h_i$ and $v_i = h_i^{-1}v_0$.

There is a slight ambiguity in what we have said about where to stop the cycle. We can either stop at the first recurrence of the vertex v_0, or we can stop at the point at which we have made a full circuit of the vertex v_0. We will discuss how to distinguish these cases shortly. If we continue until we have made a full circuit of the vertex v_0, then

$$v_0 \xrightarrow{e_0} v_1 \xrightarrow{e_1} v_2 \longrightarrow \cdots \longrightarrow v_{k-1} \xrightarrow{e_{k-1}} v_k = v_0$$

is called the **full vertex cycle** of v_0. Since we have gone all the way round the vertex, we have $e_0^{-1}\cdots e_{k-1}^{-1} = id$, that is $id = e_{k-1}\cdots e_0$. Notice that to find the relation we read off the sequence of elements e_i from *right* to *left*.

Remark 6.7. *We note that the above procedure works backwards also, in other words, given a vertex v_i, we can run the process backwards to find v_{i-1} and e_{i-1}.*

Example 6.8. Let's look at how this works in our usual concrete example, that of $\mathbb{Z}^2$ acting on $\mathbb{R}^2$ by horizontal and vertical translation by 1. This is illustrated in Figure 6.9. Using our previous notation and choosing as v_0 the top right hand vertex, we read off the cycle:

$$v_0 \xrightarrow{T^{-1}} v_1 \xrightarrow{S^{-1}} v_2 \xrightarrow{T} v_3 \xrightarrow{S} v_0$$

This gives the relation

$$id = STS^{-1}T^{-1},$$

in accordance with what we found before. You can check that $e_0 = h_1^{-1}h_0 = T^{-1}$; $e_1 = h_2^{-1}h_1 = S^{-1}T^{-1}T = S^{-1}$; $e_2 = h_3^{-1}h_2 = S^{-1}TS = T$; $e_3 = h_4^{-1}h_3 = S$ and also that $v_1 = h_1^{-1}v_0 = T^{-1}v_0$; $v_2 = h_2^{-1}v_0 = S^{-1}T^{-1}v_0 = S^{-1}v_1$; $v_3 = h_3^{-1}v_0 = S^{-1}v_0 = TS^{-1}T^{-1}v_0 = TS^{-1}v_1 = Tv_2$; $v_4 = h_4^{-1}v_0 = v_0$.

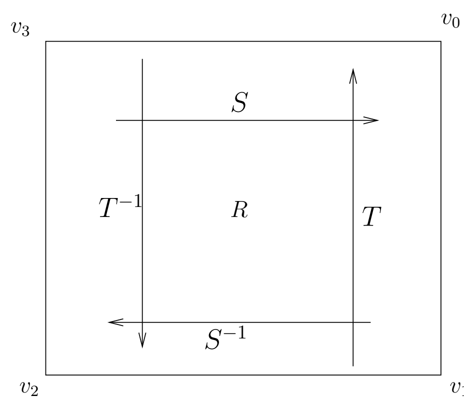

Figure 6.9: Reading off the cycle relations for $< S, T >$.

6.1.3 How do we know when to stop?

As mentioned above, there is a question about how we know when to stop the vertex cycle, either at the first recurrence of the initial vertex v_0, or when we have made a full circuit of the vertex v_0. To distinguish these cases, we look at angle sums round the vertex v_0. Let $\theta_i = \theta_{v_i}$ be the angle in R at v_i. Applying the transformation h_i (see Figure 6.8), we see that θ_i is also the angle in h_iR at v_0,see Figure 6.10. If we consider the full vertex cycle

$$v_0 \xrightarrow{e_0} v_1 \xrightarrow{e_1} v_2 \xrightarrow{e_2} \cdots v_{k-1} \xrightarrow{e_{k-1}} v_k = v_0$$

there are two possibilities: either (i) $v_r \neq v_0$ for any $r < k$; or (ii) $v_r = v_0$ for an index $0 < r < k$. We can test which case we are in by looking at the

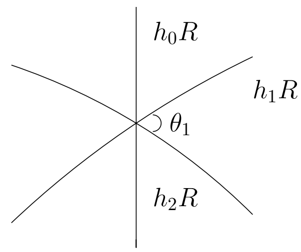

Figure 6.10: The angle at vertex h_iR at v_0 is the same as the angle in R at $h_i^{-1}v_0 = v_i$.

angle sum $\sum_{i=0}^{k-1} \theta_{v_i}$ for the vertices v_i in the cycle. In the first case, we have gone around all the copies of R which meet at the vertex v_0 and so

$$\sum_{i=0}^{k-1} \theta_i = 2\pi.$$

In the second case, we have only gone part of the way round the regions which meet at v_0 so

$$\sum_{i=0}^{k-1} \theta_i < 2\pi.$$

In fact we can say more. Since e_r is determined from v_r, the cycle repeats from v_r onwards. Thus there are, say, p repeats of the cycle before have gone all the way round R, that is $pr = k$ so that $p(\theta_0 + \cdots + \theta_{r-1}) = 2\pi$ or

$$\sum_{i=0}^{r-1} \theta_i = 2\pi/p$$

for some $p \in \mathbb{N}, p > 1$.

In both cases, we know that if we go round the full vertex cycle then $e_{k-1} \cdots e_0 = id$. In the first case, no sub-cycle $e_i^{-1} \cdots e_j^{-1}$ fixes v_0. It follows that $e_i^{-1} \cdots e_j^{-1}$ is not elliptic and $e_{k-1} \cdots e_0 = id$ gives a relation in Γ.

In the second case, we have $v_r = h_r^{-1} v_0 = v_0$. Going round p times completes the full vertex cycle so we find $(e_{k-1} \cdots e_0)^p = id$. In other words, $e_{r-1} \cdots e_0$ is elliptic with order p and fixed point v_0.

6.1.4 Examples

Let us look at some examples to see how this works. For the moment we shall assume that somehow we know that the group generated by the side pairings is discrete: the content of Poincaré's theorem which we shall below is that this is in fact *automatic* once we have checked the angle sum for each vertex cycle is correct.

The first case is illustrated by our familiar torus example. Referring back to example 6.8, the full vertex cycle is

$$v_0 \xrightarrow{T^{-1}} v_1 \xrightarrow{S^{-1}} v_2 \xrightarrow{T} v_3 \xrightarrow{S} v_0.$$

The angle in R at each vertex is $\pi/2$ and $\sum_{i=0}^{3} \theta_i = 4 \cdot \frac{\pi}{2} = 2\pi$. Thus $p = 1$ and we get the relation $STS^{-1}T^{-1} = id$ as claimed.

The second case can be illustrated by using a square in which we pair adjacent sides, as illustrated in Figure 6.11. In this case there are three vertex cycles:

$$v_2 \xrightarrow{A} v_2; \quad v_0 \xrightarrow{B} v_0; \quad v_3 \xrightarrow{B^{-1}} v_1 \xrightarrow{A^{-1}} v_3.$$

The angle at each vertex is still $\pi/2$. Thus for the cycle $v_2 \xrightarrow{A} v_2$ we have $\theta_{v_2} = \pi/2 = 2\pi/4$. We deduce that A is elliptic of order 4, that is, we have the relation $A^4 = id$. From the cycle $v_0 \xrightarrow{B} v_0$ we deduce similarly that B is elliptic of order 4 and $B^4 = id$.

Finally for the cycle $v_3 \xrightarrow{B^{-1}} v_1 \xrightarrow{A^{-1}} v_3$ we have $\theta_{v_3} + \theta_{v_1} = \pi = 2\pi/2$. We deduce the relation $(A^{-1}B^{-1})^2 = id$. In other words, BA is elliptic of order 2.

Now we look at a family of hyperbolic examples. In each case, we take a regular hyperbolic quadrilateral with angle $\alpha < \pi/2$ at each vertex. We pair opposite sides as shown in Figure 6.12. Each side is labelled with the corresponding side pairing, so that for example A takes the side with endpoints v_0, v_1 to the side with endpoints v_3, v_2.

Just as in Example 6.8, we find the cycle

$$v_0 \xrightarrow{B} v_1 \xrightarrow{A} v_2 \xrightarrow{B^{-1}} v_3 \xrightarrow{A^{-1}} v_4.$$

The angle sum round this cycle is

$$\sum_{i=0}^{3} \theta_{v_i} = 4\alpha = \frac{2\pi}{p},$$

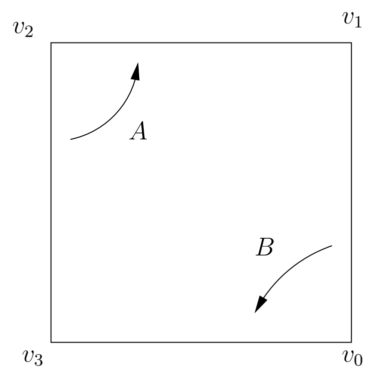

Figure 6.11: A square with different side-pairings. There are 3 vertex cycles giving relations $A^4 = B^4 = (BA)^2 = id$.

with $p \in \mathbb{N}$. From this we conclude that copies of the quadrilateral can tessellate $\mathbb{H}$ only if $\alpha = \frac{\pi}{2p}$ for some $p > 1$. In this case, $A^{-1}B^{-1}AB$ is elliptic of order p, in other words, $(A^{-1}B^{-1}AB)^p = id$. If $p = 1$, we have the Euclidean case that we have already seen.

Example 6.9. Consider the Dirichlet domain R for the group $\mathrm{SL}(2,\mathbb{Z})$ which we discussed in Section 5.3.3. Recall that R was the region bounded by the lines $\Re z = \pm 1/2$ and the circle $|z| = 1$. The vertical sides of R are paired by the map $S : z \mapsto z+1$ which is parabolic with fixed point ∞, while the finite side is matched to itself by the map $J : z \mapsto -\frac{1}{z}$ which is elliptic of order 2 with fixed point i.

Let v_0 be the bottom-right vertex $1/2 + \sqrt{3}i/2$. The vertex cycle of v_0 is

$$v_0 \xrightarrow{e_0 = S^{-1}} v_1 \xrightarrow{e_1 = J} v_2 = v_0$$

and the angle sum in v_0 is $\theta_0 + \theta_1 = \frac{2\pi}{3}$ because, as can be seen from the diagram, $\theta_0 = \theta_1 = \frac{\pi}{3}$. So we get the relation $(JS^{-1})^3 = id$.

We can check this directly. We have

$$JS^{-1} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} -1 & -1 \\ 1 & 0 \end{pmatrix}.$$

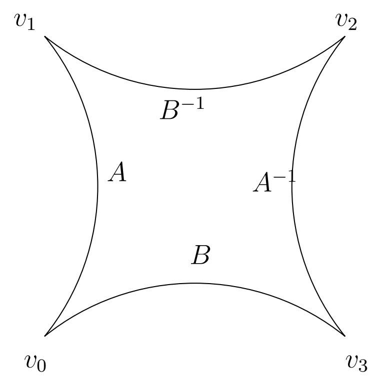

Figure 6.12: Side-pairings for a hyperbolic quadrilateral. There is one cycle with relation $(A^{-1}B^{-1}AB)^p = id$.

Thus JS^{-1} is elliptic of order 3, because

$$2\cos\phi = -1 \Rightarrow \cos\phi = -\frac{1}{2} \Rightarrow \phi = \frac{2\pi}{3},$$

where ϕ is the rotation angle round the fixed point.

To go round the full angle 2π at v_0 we need the cycle:

$$v_0 \xrightarrow{S^{-1}} v_1 \xrightarrow{J} v_0 \xrightarrow{S^{-1}} v_1 \xrightarrow{J} v_0 \xrightarrow{S^{-1}} v_1 \xrightarrow{J} v_0.$$

There is also a cycle at i, namely $i \xrightarrow{J} i$. The angle at this vertex is π, from which we deduce the relation $J^2 = id$ which we of course already know.

Hence $SL(2,\mathbb{Z})$ has two generators, S and J, satisfying the relations $(JS^{-1})^3 = id; J^2 = id$. It will follow from Poincaré's theorem (see below) that these are all the relations we need for a presentation of $SL(2,\mathbb{Z})$.

Example 6.10. Now consider the Ford domain R for $\mathrm{SL}(2,\mathbb{Z})$ that we discussed in Section 5.3.4. This is the region bounded by the vertical lines $\Re z = 0, \Re z = 1$, and the circles of radius 1 and centres $0, 1$. There are two side pairings for R:
(i) the map $\Omega : z \mapsto \frac{z-1}{z}$ which is elliptic of order 3 and which has fixed point $w_0 = e^{i\frac{\pi}{3}}$, that is $\Omega = \begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix}$;
(ii) the map $S : z \mapsto z + 1$.

If we choose the bottom-left vertex v_0, we get the vertex cycle

$$v_0 \xrightarrow{e_0=\Omega} v_1 \xrightarrow{e_1=S^{-1}} v_2 = v_0$$

and the angle sum in v_0 is $\theta_0 + \theta_1 = \pi$ because $\theta_0 = \theta_1 = \frac{\pi}{2}$. Checking by direct calculation, we have: $\Omega S^{-1} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ so $S^{-1}\Omega \neq id$, but $(S^{-1}\Omega)^2 = id$. An alternative way to see that $S^{-1}\Omega$ is elliptic of order 2 is because

$$\text{Trace } S^{-1}\Omega = 0 \Rightarrow 2\cos\phi = 0 \Rightarrow \cos\phi = 0 \Rightarrow \phi = \frac{\pi}{2},$$

where ϕ is the rotation angle round the fixed point.

Now look at the vertex $w_0 = e^{i\pi/3}$. The vertex cycle is

$$w_0 \xrightarrow{e_0=\Omega^{-1}} w_1 = v_0$$

and the angle sum in w_0 is $\theta_0 = \frac{2\pi}{3}$. So we obtain the corresponding relation $(\Omega^{-1})^3 = id$, as we already know. Poincaré's theorem will tell us that $(S^{-1}\Omega)^2 = id; (\Omega^{-1})^3 = id$, are an alternative presentation of $SL(2,\mathbb{Z})$.

In both these examples, we omitted the vertex ∞ of ∂R. In this case we have a cycle $\infty \xrightarrow{e_0=S} \infty$ and the vertex ∞ is fixed by S. There is no cycle relation, but notice that S is parabolic rather than hyperbolic. As we shall see in the next section, in fact whenever we have an ideal vertex, the group element which fixes it is necessarily parabolic.

6.1.5 Vertex cycle at ∞

The discussion up to now has looked at the pattern of copies of R round a vertex in $\mathbb{H}$. Suppose that R has a vertex on $\partial\mathbb{H}$, does this imply any new relations in G? To simplify the discussion, from now on, we will assume that R has only finitely many sides. There are two possibilities, illustrated in Figures 6.13 and 6.14.

Case (a): R has an ideal vertex on $\partial\mathbb{H}$. By this we mean that two sides of R meet at some point $v \in \partial\mathbb{H}$. An example is the Dirichlet domain for $SL(2,\mathbb{Z})$ we found in the last chapter, which has an idea vertex at ∞.

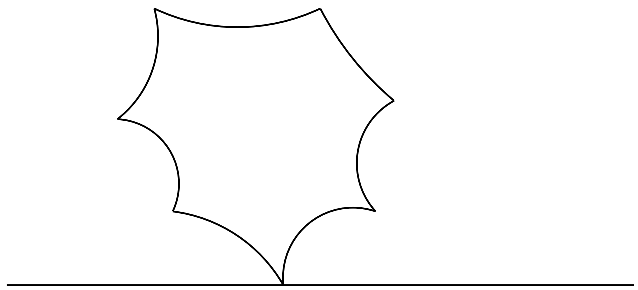

Figure 6.13: Case (a): R has an ideal vertex at $v \in \partial\mathbb{H}$.

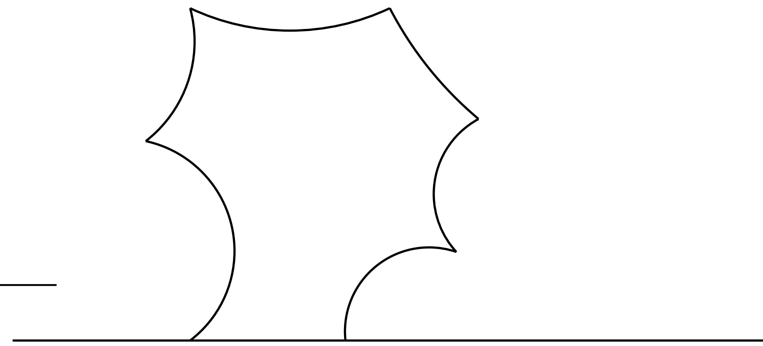

Figure 6.14: Case (b): ∂R contains a line segment in $\partial\mathbb{H}$.

Since R has only finitely many sides, it also has at most finitely many ideal vertices. Thus if we apply a side-pairing to one of the sides which ends at v, it must be mapped to another side which also ends in an ideal vertex. Thus after a finite number of applications of side-pairings, we must return to where we first started. It follows that we can associate a vertex cycle to v just as before:

$$v_0 \xrightarrow{e_0} v_1 \xrightarrow{e_1} v_2 \cdots v_{k-1} \xrightarrow{e_{k-1}} v_0 = v_k.$$

What is now the meaning of the cycle? If we consider how the copies of R are laid down adjacent to R, it is impossible to make a full circuit of v and return to R as we did in the case when the vertex was in $\mathbb{H}$. Thus in this case it is impossible that $h_k = id$. Nevertheless, after following through the cycle we have as previously $h_k^{-1} = e_{k-1} \cdots e_0$ fixes v_0. It follows that h_k must be parabolic or hyperbolic.

The following important result shows that in fact, if we assume that images of R tessellate $\mathbb{H}$, then h_k must be *parabolic*.

Proposition 6.11. *Assume that the images of R under a Fuchsian group G cover (tessellate) $\mathbb{H}$, and that this covering is locally finite. Suppose that R has an ideal vertex $v \in \partial\mathbb{H}$ with vertex cycle as above. Then $h = e_{k-1} \cdots e_0$ is parabolic.*

Proof. Suppose to the contrary that h is hyperbolic. By conjugation we may assume that $v = \infty$, so that $\mathrm{Ax}(h)$ is a vertical line with one end at ∞. Since R has an ideal vertex at ∞, we can choose a vertical line L which is eventually in R, as shown in Figure 6.15.

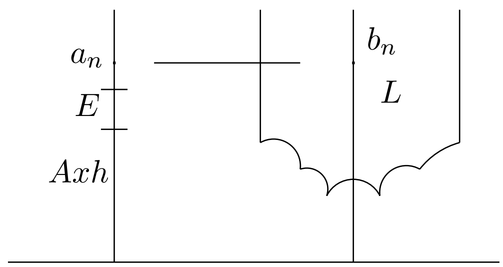

Figure 6.15: Illustration of the proof of Proposition 6.11

Choose points $b_n \in L \cap R$ such that $b_n \longrightarrow \infty$ as $n \longrightarrow \infty$ and points $a_n \in \text{Ax}(h)$ such that $\Im a_n = \Im b_n$. The lines L and $\text{Ax}(h)$ are asymptotic, so that $d(a_n, b_n) \longrightarrow 0$ as $n \longrightarrow \infty$.

Now pick a fundamental interval $E \subset \text{Ax}(h)$ for the action of h on its axis. (Any choice of interval of length equal to the translation length of h will do.) Then for all n, there exists m_n such that $h^{m_n}(a_n) \in E$. Since $a_n \longrightarrow \infty$, an infinite number of these m_n must be distinct. Since $d(a_n, b_n) \longrightarrow 0$ we have $d(h^{m_n}a_n, h^{m_n}b_n) \longrightarrow 0$ and hence $h^{m_n}(b_n) \longrightarrow E$. But $h^{m_n}(b_n) \in h^{m_n}R$, so that this shows that an infinite number of copies of R accumulate on E, which is impossible by the assumption of local finiteness of the tessellation of $\mathbb{H}$ by copies of R. So h must be parabolic as claimed. □

Case (b): $\partial\overline{R}$ contains an interval in $\partial\mathbb{H}$, precisely, R has two infinite sides whose endpoints bound an interval on $\partial\mathbb{H}$.

This is illustrated in Figure 6.14. We can apply the side pairings in the same pattern as previously but now the image regions $h_0R, h_iR, \ldots h_kR$ are adjacent one to the next, each containing intervals which are adjacent along I, see Figure 6.16. Let I be the connected component of $\bar{R} \cap \partial\mathbb{H}$ in question, so that I is an interval in $\partial\mathbb{H}$. This time we argue that the chain of adjacent intervals $I = I_0, I_1, \ldots, I_n$ which appear on $\partial\mathbb{H}$ as edges of the regions $R = h_0R, h_1R \ldots$ must eventually return to an interval $I_k = h_kI_0$ for some cycle $h_k^{-1} = e_{k-1} \cdots e_0$. Let $h = h_k$.

Lemma 6.12. *Let I be a connected component of $\bar{R} \cap \partial\mathbb{H}$ (so I is a closed subinterval of $\partial\mathbb{H}$. Then I contains no parabolic or hyperbolic fixed points.*

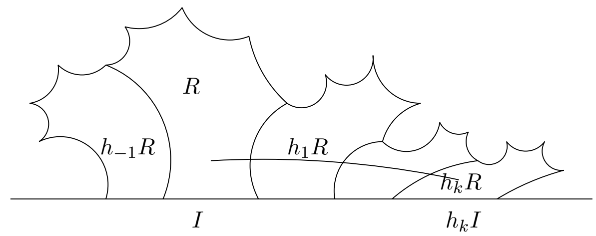

Figure 6.16: Images of the fundamental domain R under a vertex cycle in case (b).

Proof. First we let $J = \text{Int}I$, so that J is an open interval in $\partial\mathbb{H}$. We first show that J contains no fixed points of G. Suppose to the contrary that $z \in J$ is the attracting fixed point of some $g \in G$. Since R has only finitely many sides, there is a neighbourhood U of z in $\mathbb{H} \cup \partial\mathbb{H}$ so that $U \cap \partial\mathbb{H} \subset J$ and $U \cap \mathbb{H} \subset R$. Pick $w \in J$ close to z. The images $g^n(w)$ accumulate on z and hence eventually lie in J. Since $g^n(w) \in g^n\bar{R}$, it follows that there are points in $g^n(R)$ which eventually lie in U. But this is impossible, since $g(R) \cap R = \emptyset$ for $g \neq id$.

Finally we have to deal with the endpoints of I. Applying the appropriate side-pairings, we see that on one side, I is adjacent to the next interval I_1 in the cycle. On its other side and running the cycle backwards, it is adjacent to a similar interval I_{-1}. By the same argument as above, the open intervals $\text{Int}I_1$ and $\text{Int}I_{-1}$ contain no fixed points. Now the argument above applied to $I \cup I_{-1} \cup I_1$ shows that the endpoints of I are also not fixed points. □

Proposition 6.13. *Assume that the images of R under a Fuchsian group G tessellate $\mathbb{H}$, and that this covering is locally finite. If $\partial\overline{R}$ contains an interval in $\partial\mathbb{H}$, then either the corresponding vertex cycle h is hyperbolic or G is elementary.*

Proof. There is now reason why we cannot continue the cycle at I indefinitely in both directions, producing a sequence of adjacent intervals
$\dots, I_{-2}, I_{-1}, I, I_1, I_, \dots$ By Lemma 6.12, there are no fixed points of G in $K := \bigcup_{-\infty}^{\infty} I_n$.

Now there are three possibilities:
(i) $K = \partial\mathbb{H}$. In this case G can have no fixed points on $\partial\mathbb{H}$. Thus all elements

of G are elliptic, which as we have seen forces G to be elementary, generated by a finite order elliptic element. It is not hard to argue that the elliptic generator is h and that the images of $\cup_{i=0}^{k-1} I_i$ cover $\partial\mathbb{H}$.
(ii) $K = \partial\mathbb{H} \setminus \{x_0\}$ for some $x_0 \in \partial\mathbb{H}$. In this case G can have at most one fixed point on $\partial\mathbb{H}$, on which images of $\cup_{i=0}^{k-1} I_i$ accumulate on x_0 from both sides. It is not hard to see that the only option is that h is parabolic with fixed point x_0. In this case also, G is elementary generated by h.
(iii) $K = \partial\mathbb{H} \setminus L$ for some non-empty interval L. In this case the images under h and h^{-1} of $\cup_{i=0}^{k-1} I_i$ accumulate on the two endpoints of L. Thus the two endpoints of L are fixed points of h, so h is hyperbolic.

□

6.1.6 Poincaré's Theorem

In the above sections we have shown that of R is a convex polygon which is a locally finite fundamental domain for a discrete group G acting on $\mathbb{H}$, then:
(a) the angle sum corresponding to each vertex cycle with corresponding vertex $v \in \mathbb{H}$ is $\sum_{i=0}^{k-1} \theta_i = \frac{2\pi}{p}$ with $p \in \mathbb{N}$ and:
• if $p = 1$, then we get a corresponding relation $e_{k-1} \cdots e_0 = id$;
• if $p > 1$, then $e_{k-1} \cdots e_0$ is elliptic of order p, so that $(e_{k-1} \cdots e_0)^p = id$.
(b) if R is finite sided and has an ideal vertex $v \in \partial\mathbb{H}$, then the vertex cycle corresponding to v is parabolic.

A famous theorem due to Poincaré says that we can reverse these conditions:

Theorem 6.14 (Poincaré's theorem)**.** *Suppose that $R \subset \mathbb{H}$ is finite sided convex polygon whose sides are identified in pairings by isometries $\{g_1, \dots, g_r\} = G^*$. Say the above conditions (a) and (b) hold. Then:*
(i) G^* generates a Fuchsian group G;
(ii) R is a locally finite fundamental domain for the G action;
(iii) $G = < g_1, \dots, g_r :$ vertex relations $>$*, that is the vertex relations are all you need!*

Notice that there is *no* requirement on cycles associated to sides of $\bar{R}$ contained in $\partial\mathbb{H}$. If such sides exist, then either G is elementary or the cycle is automatically hyperbolic. The proof of this theorem is highly non-trivial and will be the topic of Chapter 7.

The theorem allows us finally to produce large numbers of interesting examples of Fuchsian groups. All you have to do is find a polygon R and a suitable set of side pairings. There is enormous freedom in how you can do this. Below, we give just a few examples, for more see Example Sheet 8.

Remark 6.15. If condition (a) fails, so the copies of R round a finite vertex don't fit together properly, then it is not hard to see that the group $G = < G^* >$ can't possibly be Fuchsian. However, it is possible that G may be Fuchsian while (b) fails, when the cycle associated to the ideal vertex is hyperbolic not parabolic. In this case, what happens is that even though G is discrete, the images of $\bar{R}$ do not cover all of $\mathbb{H}$ and moreover their union is not locally finite, in other words, their images may accumulate on a line which is itself not in the union of their images. This situation is really rather subtle. We shall discuss such issues further in Chapter 7.

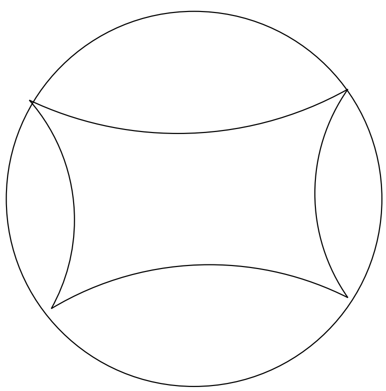

Figure 6.17: Four sided ideal quadrilateral with opposite sides paired. In this case the commutator is parabolic.

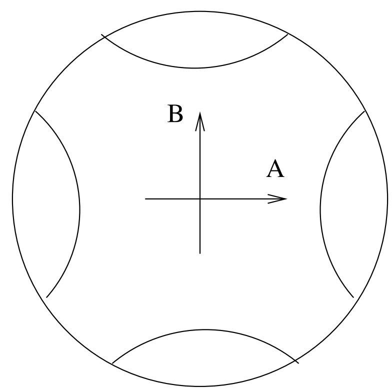

Figure 6.18: Four sided region whose boundary contains intervals on $\partial\mathbb{H}$. In this case the commutator is hyperbolic.

Example 6.16. We have already studied the example of a four sided fundamental domain with opposite sides paired by transformations A, B, see Figure 6.12. In Figures 6.17 and 6.18 we have also four sided regions R with opposite sides paired; in the first case the four sides meet in four ideal vertices and in the second ∂R includes 4 intervals on $\partial\mathbb{H}$. In all cases, we have the same vertex cycle $A^{-1}B^{-1}AB$. In the first case, for images of R to tesselate, then we can conclude that $A^{-1}B^{-1}AB$ must be parabolic. In the second situation, it is automatically hyperbolic. It is possible to find side pairings which match the sides in Figure 6.17 such that the commutator is hyperbolic, but the images of R in this case do not tesselate $\mathbb{H}$, see Chapter 7.

Example 6.17. The regular octagon. Now we look at two examples in which R is a regular hyperbolic octagon with vertex angle $\pi/4$. First consider the set of side pairings shown in Figure 6.19. In this case the vertex cycle is

$$v_0 \xrightarrow{B} v_1 \xrightarrow{A^{-1}} v_2 \xrightarrow{B^{-1}} v_3 \xrightarrow{C} v_4 \xrightarrow{D} v_5 \xrightarrow{C^{-1}} v_6 \xrightarrow{D^{-1}} v_7 \xrightarrow{A} v_8 = v_0.$$

Since there are eight terms in the cycle and the angle at each vertex is $\pi/4$, we find the total angle sum is $8 \times \pi/2 = 2\pi$. Thus the relation in this case is

$$AD^{-1}C^{-1}DCB^{-1}A^{-1}B = id.$$

Cyclically permuting and inverting, this becomes

$$A^{-1}B^{-1}ABC^{-1}D^{-1}CD = id.$$

Note that the order of terms in this relation is *not* the same as the order of labels round ∂R.

By Poincaré's theorem, the group generated by A, B, C, D is Fuchsian and the images of the octagon under G tesselate $\mathbb{H}$.

If we change the arrangement of the side pairings around ∂R, everything changes. Figure 6.20 shows the case in which opposite sides are paired. In this case the cycle sequence is:

$$v_0 \xrightarrow{D^{-1}} v_1 \xrightarrow{C} v_2 \xrightarrow{B^{-1}} v_3 \xrightarrow{A} v_4 \xrightarrow{D} v_5 \xrightarrow{C^{-1}} v_6 \xrightarrow{B} v_7 \xrightarrow{A^{-1}} v_8 = v_0.$$

So the relation in this case is

$$A^{-1}BC^{-1}DAB^{-1}CD^{-1} = id.$$

Example 6.18. (Schottky groups)
Let $C_1, \ldots, C_{2k}$ be pairwise disjoint circles arranged in such a way that if D_i is the disc inside C_i, then $D_i \cap D_j = \varnothing$ when $i \neq j$. Suppose they are paired by isometries $g_1, \ldots, g_k$, in such a way that $g_i(C_i) = C_{i+k}, i = 1, \ldots, k$ and so that $g_i(D_i) = \mathbb{H} \setminus D_{i+k}$. Then:
(i) $G =< g_1, \ldots, g_k >$ is a free group;
(ii) the region R outside all the discs D_i, for $i = 1, \ldots, 2k$, is a fundamental domain for G.
A group generated in this way is called **Schottky group**.

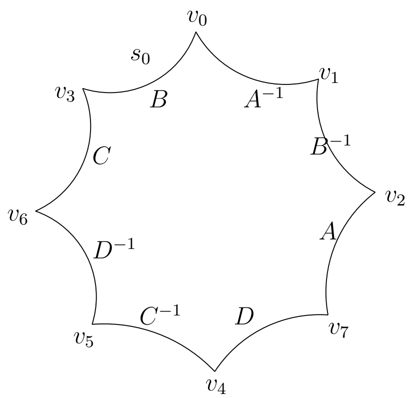

Figure 6.19: Side pairings of a regular octagon giving the relation $D^{-1}C^{-1}DCB^{-1}A^{-1}BA = id$.

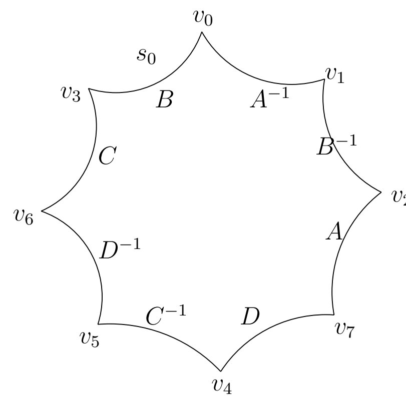

Figure 6.20: Side pairings of a regular octagon matching opposite sides. The relation is $A^{-1}BC^{-1}DAB^{-1}CD^{-1} = id$.

Proposition 6.19. *If G is not elementary and discrete, it contains a free group of rank 2.*

Proof. First consider a single hyperbolic isometry with axis L and translation length d. Let C, C' which meet L orthogonally at distance d apart, and let D, D' be the disjoint disks bounded by C, C'. Then $h(C) = C'$ and $h(D) = \mathbb{H} \setminus C'$.

Now pick 2 hyperbolic isometries g_1 and g_2 with distinct axes L_1, L_2 and translation lengths d_1, d_2. We want to make the above construction for each axis separately, but we want to do it in such a way that the four circles C_i, C_i are disjoint, $i = 1, 2$. This may not be possible. However, if we take $N \in \mathbb{N}$ sufficiently large, then the 4 circles orthogonal to L_1, L_2 at distances Nd_1 and Nd_2 apart are disjoint. So picking sufficiently high powers of g, g' we can arrange a fundamental domain for a Schottky group as above. In fact this is just the 4 sided quadrilateral example illustrated in Figure 6.18. $\square$

Chapter 7

Hyperbolic structures on surfaces

7.1 Hyperbolic structures on surfaces

Definition 7.1. *A* ***hyperbolic structure*** *on a topological surface S is a maximal collection of coordinate charts, that is open sets $U_i \subset S$ and maps $\Phi_i : U_i \longrightarrow \mathbb{H}$ such that:*
(i) $\Phi_i : U_i \longrightarrow \Phi_i(U_i)$ is a homeomorphism:
(i) the sets U_i cover S;
(ii) the "overlap maps" are isometries, that is

$$\gamma_{ij} = \Phi_i \Phi_j^{-1} : \Phi_j(U_i \cap U_j) \longrightarrow \Phi_i(U_i \cap U_j) \in \operatorname{Isom} \mathbb{H}.$$

We call a surface with a hyperbolic structure, a **hyperbolic surface**. The definition of a hyperbolic structure is just like the definition of a manifold, except that the coordinate charts map to $\mathbb{H}$ and the overlap maps have to be hyperbolic isometries. Of course one can make a much more general definitions of a geometric structure on a manifold, for example see the interesting article [17].

We have available two slightly different ways of constructing a hyperbolic structure on a surface. The first method requires only a polygon R together with a set of side pairing isometries, and *does not* require a Fuchsian group G for which R is a fundamental domain. In contrast, the second requires a Fuchsian group G acting freely on $\mathbb{H}$, but *does not* depend on the choice of a fundamental domain.

Method 1.
Suppose given a hyperbolic polygon, and suppose that the sides of R are paired by a set of hyperbolic isometries G_0 as discussed above. We can form a surface by identifying two point if they are matched by one of these isometries, more precisely, we define an equivalence relation $\sim$ on $\overline{R}$ by:

$$x \sim y, x, y \in \overline{R} \text{ iff } x = g(y) \text{ for some } g \in G_0.$$

It is clear that $R/\sim$ is a surface, which we can topologise with the quotient topology. Points in the interior of R are in one point equivalence classes while points in the interior of a side of $\partial\mathbb{R}$ are in two point classes. If x is a vertex, then its equivalence class consists exactly of all the vertices in its vertex cycle.

If the vertex cycles do not give any elliptic elements, then the surface $R/\sim$ carries a natural hyperbolic structure.

To see this, we have to define the coordinate charts. Think of R as embedded in $\mathbb{H}$ by the inclusion $R \hookrightarrow \mathbb{H}$.
(i) If $x \in \text{Int}(R)$, then $\Phi := id$.
(ii) If $z \in s$ for some side s of R, but z is not a vertex, let $R, g(R)$ be the copies of R which are adjacent along s. (So that s is associated to the side pairing g^{-1}.) Choose $r > 0$ such that $B_r(z) \cap R$ is an open half disc centred on z and bounded by s, while $B_r(g^{-1}z) \cap R$ is a similar half disc centred on $g^{-1}(z)$. (This is possible since z is not a vertex of R and R is a convex polygon.) Let $A = (B_r(z) \cap R) \cup (B_r(z) \cap s)$ (so A is an open half disc in R, together with an open segment in s centred at z) and likewise let $B = (B_r(g^{-1}z)) \cap R) \cup (B_r(g^{-1}z)) \cap g^{-1}s)$. In the quotient $R/\sim$, $x \in A \cap s$ is identified to $y \in B \cap s$ exactly when $x = g(y)$; in other words, A and B are glued along $s \cap A$ and $g^{-1}(s) \cap B$.

Let U be the projection to $R/\sim$ of $A \cup B$. Then U is an open neighbourhood of the image $\bar{z}$ of z in $R/\sim$. We define the associated chart $\Phi : U \longrightarrow \mathbb{H}$ as follows:
• If $x \in A$, then $\Phi(x) := x$.
• If $x \in B$, then $\Phi(x) := g(x)$.
(iii) If z is a vertex of R, we have to make a chart by gluing together all the copies of R which make a cycle round the vertex z. The details are an exercise.

Exercise 7.2. *Verify that these charts satisfy the definition of a geometric structure above.*

For future use, we observe that we can define a metric on $R/\sim$ as follows:

$$\bar{d}([x],[y]) = \inf \sum_{i=0}^{n} d(x_i, x_i')$$

where $x_0 \in [x], x_n' \in [y]$ and $x_i' \sim x_{i+1}$ for $i = 0, \ldots, n-1$.
The symmetry of $\bar{d}$ and the triangle inequality are obvious. Thus to prove that $\bar{d}$ defines a metric, we have to prove only that

$$\bar{d}([x],[y]) = 0 \Rightarrow [x] = [y.]$$

This can be done by noting that every point is contained in a ball in which the restricted metric is hyperbolic.

Method 2. Suppose that G is a Fuchsian group acting freely on $\mathbb{H}$. (To act freely means that $gx = x$ implies that $g = id$ for all $x \in \mathbb{H}$.) Let $\mathbb{H}/G$ denote the space of G-orbits. There is a natural projection $\mathbb{H} \longrightarrow \mathbb{H}/G$; we give $\mathbb{H}/G$ the quotient topology. One can show (exercise!) that $\mathbb{H}/G$ is Hausdorff. A similar definition to $\bar{d}$ above makes the quotient $\mathbb{H}/G$ a metric space.

Each point in $[x] \in \mathbb{H}/G$ is a G-orbit in $\mathbb{H}$. To define the hyperbolic structure, pick $x \in \mathbb{H}$ which projects in $[x]$. Choose a open ball $B_r(x)$ containing x in $\mathbb{H}$ which contains no other orbit points (possible since G acts freely and properly discontinuously). Then $B_r(x)$ projects to a set $U \subset \mathbb{H}/G$ whose lift $\tilde{U}$ to $\mathbb{H}$ consists of the disjoint open balls $B_r(gx), g \in G$. This set is open in $\mathbb{H}$ so that U is open in $\mathbb{H}/G$. We can define a chart as the map which identifies U with the ball $B_r(x) \subset \mathbb{H}$.

Exercise 7.3. *Check that this defines a geometric structure on* $\mathbb{H}/G$.

Exercise 7.4. *Check that if R is a fundamental domain for a Fuchsian group G acting on $\mathbb{H}$ then the two ways of forming a quotient give the same hyperbolic structure.*

If G contains elliptic elements, then it does not act freely on $\mathbb{H}$. The problem in extending the above ideas to this case is that in the neighbourhood of an elliptic fixed point, there will never be a chart which maps *bijectively* to $\mathbb{H}$. Instead, there is a chart which maps homeomorphically to a neighbourhood of the singular point in the quotient $\mathbb{H}/ < T >$, where T is a finite order elliptic. Such a structure is called an *orbifold*.

Definition 7.5. *A surface S is a hyperbolic* ***orbifold*** *if it satisfies the same conditions as that of a geometric structure above, except that at finitely many points, the chart maps to the quotient $U/ < T >$ where U is a neigborhood of an elliptic fixed point v and T is an element of finite order fixing v.*

Groups which contain no finite order elements (in our case no elliptics) and called *torsion free*.

Definition 7.6. *A hyperbolic surface or orbifold S is said to be* complete *if the metric $\bar{d}$ above is metrically complete, in other words, every sequence of points $x_n \in S$ which is $\bar{d}$ Cauchy, has a subsequence which converges to a point in S.*

The importance of this definition will be apparent later.

7.2 Cusps, funnels and cone points

If the boundary of the fundamental region or polygon R meets $\partial\mathbb{H}$, the glued up surface $S = R/\sim$ has boundary. Likewise if the group G contains elliptic elements, S contains some distinguished points. These are features of the surface which we can describe in more detail as follows.

As we have already seen, elliptic fixed points correspond to points where S is locally like $\mathbb{H}/ < T >$, where T is elliptic. These are called **cone points** of S. In a neighbourhood of such a point, S looks like a cone with angle $2\pi/n$ where n is the order of the elliptic at that point, see Figure 7.3.

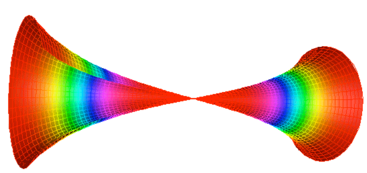

Figure 7.1: Cone points. This picture is slightly misleading as on a surface with a hyperbolic structure the cone points would not meet like this.

To deal with the ends of S, first look at two types of hyperbolic structures on annuli:
(a) $\mathbb{H}/ < T >$, where T is hyperbolic.

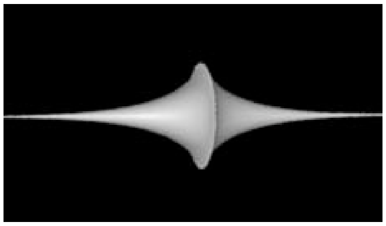

Figure 7.2: Back to back cusps. This picture is also slightly misleading as on a surface with a hyperbolic structure would not contain a singular disc. Such surfaces cannot be isometrically emebdded into in Euclidean space $\mathbb{R}^3$.

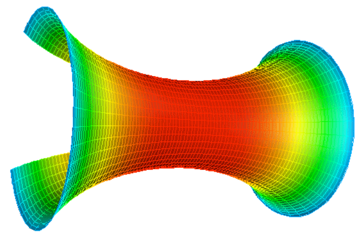

Figure 7.3: Hyperbolic annulus. A funnel on a surface looks like one end of this cylinder.

(b) $\mathbb{H}/ < T >$, where T is parabolic.

We say that S has a **cusp** if there is a neighborhood of a component of ∂S which has a chart to a horoball neighborhood of a standard cusp, see Figure 7.2.

We say that S has a **funnel** if there is a neighborhood of a component of ∂S which has a chart to an end of a hyperbolic annulus, see Figure 7.3.

Note that these elliptic (or parabolic) hyperbolic surfaces correspond to the 3 surfaces of constant negative curvature of revolution, see in [18], p. 242.

Notice that in all these cases, the quotient S with the induced metric $\bar{d}$ is metrically complete. We shall study a more subtle case in which S is not metrically complete in the next section.

It is not hard to check the following Lemma.

Lemma 7.7. *Parabolic (or elliptic) elements of G correspond precisely to parabolic points of $\mathbb{D}/G$. The number of such points on $\mathbb{D}/G$ is equal to the number of conjugacy classes of maximal parabolic (or elliptic) elements of G.*

7.3 Incomplete hyperbolic structures

Suppose we have a convex finite sided polygon R and associated side pairings, and suppose there are cusp vertices on $\partial\mathbb{H}$ for which the associated vertex cycle is *hyperbolic*. As we shall see, this does not necessarily imply that the group generated by the side pairings is not discrete; it does however mean that the tessellation by images of R is no longer locally finite and does not cover the whole of $\mathbb{H}$.

The pictures below, taken from [16], help us to understand how this can come about. We want to study a region R which has a cusp vertex fixed by a single hyperbolic transformation h. In Figure 7.4, the heavy line $\Re z = a_\infty$ is the axis of a hyperbolic element $h(z)$ with attracting fixed point a_∞ and repelling fixed point ∞. Thus h is a contraction (homothety) by a factor $\lambda < 1$; precisely $h(z) = a_\infty + \lambda(z - a_\infty)$. The lines $L_n := \{\Re z = a_n\}$ are successive images of $L_0 = \{\Re z = 0\}$ under powers of h; thus $a_{n+1} = h(a_n) = h^n(0)$ and $h(L_n) = L_{n+1} = h^{n+1}L_0$.

Now imagine that the cusp vertex of R is the point ∞ and that the parts of L_0 and L_1 above some definite height are the sides of R which meet at ∞. Suppose that the vertex cycle at ∞ is just the hyperbolic element h. Take a sequence of points $z_n \in L_n$ on the same horizontal level, ie such that $\Im z_n = \Im z_0$ for all n. Each z_n is carried back under h^{-n} to a point in L_0. It is easy to see that $d_{\mathbb{H}}(z_n, z_{n+1})$ decreases geometrically, so (z_n) is a Cauchy sequence in $\mathbb{H}$. The limit is clearly the point $a_\infty + i\Im z_0$ at the same horizontal level as z_0 on $\mathrm{Ax}\,h$. Let $[z_n]$ be the image of z_n on $S = R/\sim$. Since the projection to S clearly contracts distance, the points $[z_n]$ are a Cauchy sequence on S which has no limit in S. In other words, S is not metrically complete.

An alternative view of this phenomenon is that the geodesic lines shown on the left in Figure 7.5 do not extend indefinitely, rather they come to an end when they reach $\mathrm{Ax}\,h$.

The right hand frame in Figure 7.5 shows the quotient of a neighbourhood of $\mathrm{Ax}\,h$ by the action of h. The black boundary curve is exactly the length of the translation length of h, and is missing from the quotient surface S. The image of the line L_0 (which in the quotient is equivalent to every line L_n) is seen spiralling in towards the missing boundary.

Finally, let us see this phenomenon at work in an actual Fuchsian group. Figure 7.6 shows the tessellation of $\mathbb{H}$ by the group generated by the two

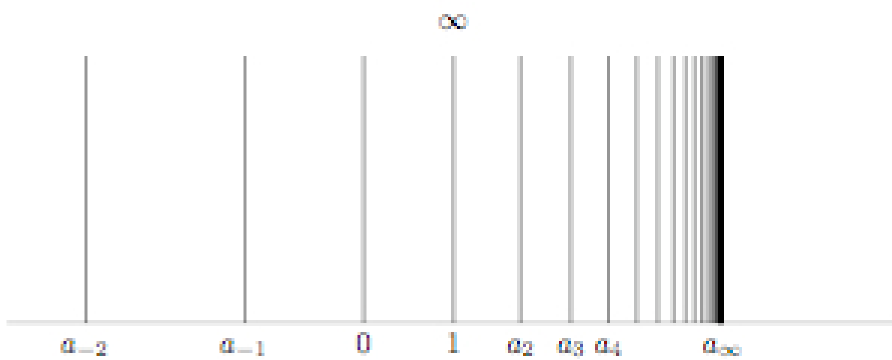

Figure 7.4: Accumulation of strips under iteration of a hyperbolic element h with fixed point at a_∞.

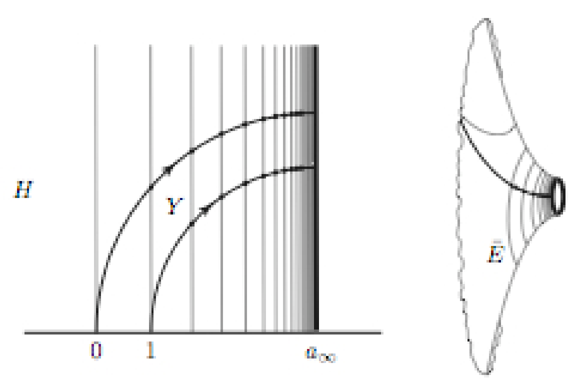

Figure 7.5: The quotient of the previous figure by hyperbolic h is a cylinder. The dark boundary is missing from the tessellation by images of the strip.

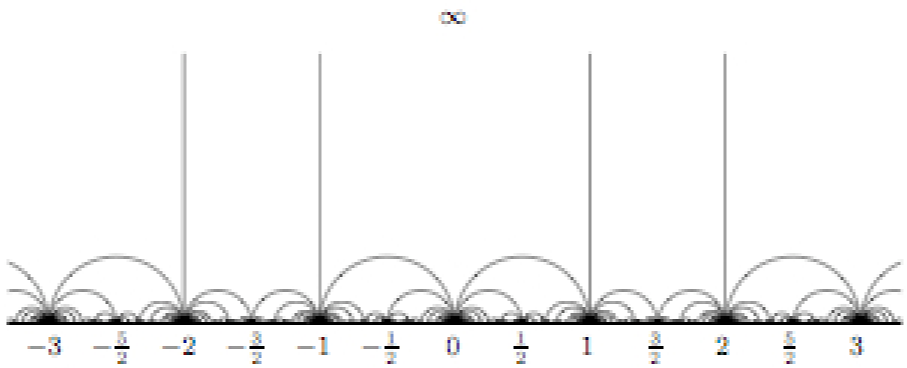

Figure 7.6: Tesselation of $\mathbb{H}$ by a once punctured torus group with parabolic commutator.

transofrmations $A = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$ and $B = \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}$. A fundamental domain for the corresponding group of isometries (which has index six in $PSL(2, \mathbb{Z})$) is the region bounded by the lines $\Re z = \pm 1$ and the semicircles from $\infty 1$ to 0. You can check that A maps $[-1, \infty]$ to $[0, 1]$ and B maps $[1, \infty]$ to $[0, -1]$. This region is an ideal quadrilateral and opposite sides are matched in pairs. You can also check that the vertex cycle corresponding to ∞ is the commutator $B^{-1}A^{-1}BA$ which one can verify by direct computation is the parabolic transformation $z \mapsto z + 6$. Thus in this case the region R with its side pairings A, B satisfies the conditions of Poincaré's theorem, and the images of R indeed forma a locally finite tessellation of $\mathbb{H}$.

Figure 7.7 shows a very similar set-up. We start with the same region R and the sides are still paired by hyperbolic isometries A' and B' say. However now A', B' are chosen such that their commutator $B'^{-1}A'^{-1}B'A'$ is *hyperbolic*.[1] The result is that instead of tessellating the whole plane, the images of R under $G =< A', B' >$ accumulate on every image of the axis of the commutator, shown by heavy black lines in the figure. In this situation, S is not metrically complete; geodesics on S do not extend indefinitely; the

[1] Note that A and A' both pair the same pair of sides $[\infty, -1]$ and $[0, 1]$. If these lines had finite endpoints in $\mathbb{H}$, the choice of pairing transformation would be unique. However since the sides in question are geodesics with endpoints on $\partial\mathbb{H}$, there is a one-parameter family of freedom in the choice of pairing. This is what makes it possible to choose A' as required.

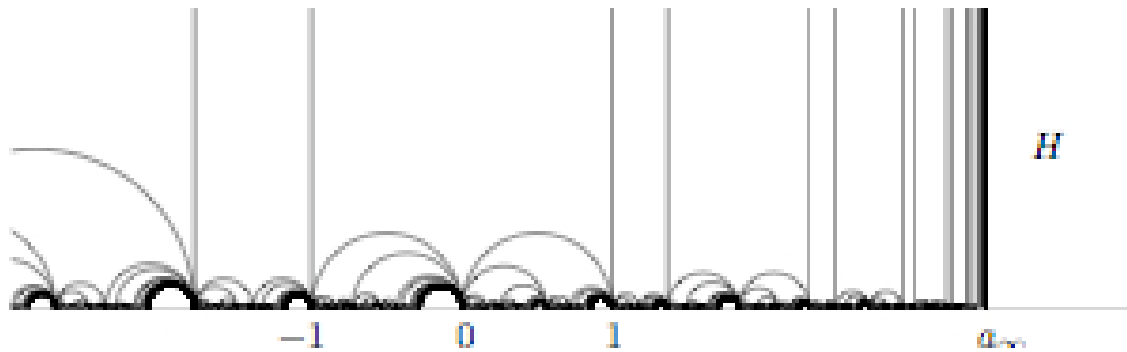

Figure 7.7: Tessellation of $\mathbb{H}$ under a two generator group which generate a once punctured torus group with hyperbolic commutator.

group $G = < A', B >$ (ie the group generated by the the side pairings of R) is discrete; the images of R are mutually disjoint but do not cover all of $\mathbb{H}$; and finally the images are not locally finite, since an infinite number of images accumulate in a neighbourhood of each image of Ax h.

It is this situation which is being avoided by the three equivalent conditions in Theorem 7.23 below. For further discussion and other pictures, see [14].

7.4 Covering spaces

In this section, we sketch briefly what is needed from the theory of covering spaces to understand Poincaré's theorem and its proof. All the concepts discussed in this section are in Introduction to Topology course MA 3F1 and can be found in many books on topology, for example [19] p. 227-232 and [20] Section 1.3.

All the spaces below are topological spaces and 'map' means 'continuous map'.

Definition 7.8. *The map $f : Y \longrightarrow X$ is a **covering map** if f is surjective and if each $x \in X$ has a neighborhood U which is evenly covered, that is $f^{-1}(U) = \sqcup V_i$ (disjoint union) such that $f|_{V_i}$ is a homeomorphism to U. The space Y is called a **covering space** of X.*

Definition 7.9. *If we have a covering map $f : Y \longrightarrow X$ and another space*

Z with a map $g : Z \longrightarrow X$, then a map $\tilde{g} : Z \longrightarrow Y$ such that $f \circ \tilde{g} = g$ is called a ***lift*** *of the map g.*

Theorem 7.10. *(Path lifting) Suppose $f : Y \longrightarrow X$ is a covering space. Suppose $\alpha : [0,1] \longrightarrow X$ is a path with $\alpha(0) = x_0$, and suppose $y_0 \in f^{-1}(x_0)$. Then there exists a unique path $\tilde{\alpha} : [0,1] \longrightarrow Y$ such that such that $f \circ \tilde{\alpha} = \alpha$ and such that $\tilde{\alpha}(0) = y_0$, in other words, such that the path $\tilde{\alpha}$ is a* lift *of α*

7.4.1 Fundamental group

Let X be a topological space. Recall that two paths $\alpha_1 : [0,1] \longrightarrow X$ and $\alpha_2 : [0,1] \longrightarrow X$ with $\alpha_1(0) = \alpha_2(0)$ and $\alpha_1(1) = \alpha_2(1)$ are called **homotopic** if you can continuously deform one to the other. More formally (see, for example, [19]):

Definition 7.11. *Two paths $\alpha_1 : [0,1] \longrightarrow X$ and $\alpha_2 : [0,1] \longrightarrow X$ with $\alpha_1(0) = \alpha_2(0) = p$ and $\alpha_1(1) = \alpha_2(1) = p'$ are* ***homotopic*** *if there exists a continuous map $F : [0,1] \times [0,1] \longrightarrow X$ such that $\alpha_1(t) = F(t,0)$, $\alpha_2(t) = F(t,1)$, $F(0,s) = p$ and $F(1,s) = p'$ for all $(t,s) \in [0,1] \times [0,1]$.*

Theorem 7.12. *(Homotopy lifting) Let $f : Y \longrightarrow X$ be a covering map and let $F : [0,1] \times [0,1] \longrightarrow X$ be a homotopy between two paths α_1 and α_2 in X with $\alpha_1(0) = \alpha_2(0) = p$ and $\alpha_1(1) = \alpha_2(1) = p'$. Let $\tilde{\alpha}_1$ and $\tilde{\alpha}_2$ be liftings of α_1 and α_2 respectively. Then there exists a unique homotopy $\widetilde{F} : [0,1] \times [0,1] \longrightarrow Y$ between $\tilde{\alpha}_1$ and $\tilde{\alpha}_2$ such that $\widetilde{F}$ lifts F.*

The set of the homotopy classes of loops based at $x_0 \in X$ is called the **fundamental group** of X and is denoted $\pi_1(X, y_0)$. It is a group under the composition of paths.

Corollary 7.13. *If $f : Y \longrightarrow X$ is a covering map, then f induces an injection of the fundamental groups $f_* : \pi_1(Y, y_0) \longrightarrow \pi_1(X, x_0)$.*

Proof. We assume that $f(y_0) = x_0$. Suppose that β is a loop in Y based at y_0 which maps to the trivial loop in X. This means that there is a homotopy $F : [0,1] \times [0,1] \longrightarrow X$ between $f \circ \beta$ and the trivial path $g(t) \equiv x_0$. the path $g(t) \equiv x_0$ clearly lifts to the trivial path $\tilde{g}(t) \equiv y_0$. and the homotopy F lifts to a homotopy between $\tilde{g}$ and β. □

7.4.2 The Universal Covering Space

Definition 7.14. *A space Y is* ***simply connected*** *if every loop in Y is homotopic to a point. In particular, $\pi_1(Y, y_0) = \{\text{id}\}$.*

Given a space X, the *universal cover* of X is a simply connected covering space of X. Subject to mild hypotheses on X, such a space always exists and is unique up to homeomorphism. In more detail, fix a point $x_0 \in X$ and let

$$\widetilde{X} = \{\text{homotopy classes of paths } \alpha : [0, 1] \longrightarrow X, \alpha(0) = x_0\}.$$

With a suitable topology on $\tilde{X}$ (roughly speaking, the topology obtained by lifting small neighbourhoods in X to $\tilde{X}$, see for example [19] Theorem 10.19):

Theorem 7.15. *Under mild hypotheses on X explained below, the map $p : \widetilde{X} \longrightarrow X$, defined by $p(\alpha) = \alpha(1)$, is a covering map. Moreover $\widetilde{X}$ is simply connected. If $p' : \widetilde{X}' \longrightarrow X$ is any other simply connected covering space of X then there exists a homeomorphism $h : \widetilde{X} \longrightarrow \widetilde{X}'$ such that $p' \circ h = p$.*

The mild hypotheses on X are that it must be:
(i) connected
(ii) locally path-connected, that is, each point $x \in X$ has a neighbourhood base of path connected neighbourhoods;
(iii) semi-locally simply connected, that is, each point $x \in X$ has a neighbourhood in which every loop is null homotopic.

The space $\widetilde{X}$ is called the **universal cover** of X. The canonical example to focus on here is:

Example 7.16. *Let G be a discrete group acting freely on $\mathbb{H}$ and let $X = \mathbb{H}/G$. Then $p : \mathbb{H} \longrightarrow \mathbb{H}/G$ is a covering map. Since $\mathbb{H}$ is simply connected, it can be identified with the universal cover of X.*

A map $g : \widetilde{X} \longrightarrow \widetilde{X}$ which commutes with $p : \widetilde{X} \longrightarrow X$ is called **covering transformation**. The set of covering transformations form a group, the covering group of X. It is a standard result that the covering group can be identified with the fundamental group $\pi_1(X)$.

In our example $X = \mathbb{H}/G$, where G is a torsion free Fuchsian group, this identification can be seen very concretely as follows. Each $g \in G$ defines an isometry $\mathbb{H} \longrightarrow \mathbb{H}$ which commutes with the projection to X. It is not hard to see that *every* map $\mathbb{H} \longrightarrow \mathbb{H}$ which respects G-orbits must have this form. Thus we can identify G with the group G of covering transformations of X. In particular, fixing a point $\tilde{x}_0 \in p^{-1}(x_0)$, then for every $g \in \Gamma$ we have a corresponding point $g\tilde{x}_0 \in \mathbb{H}$. Observing that the path from x_0 to $g\tilde{x}_0$ projects to a loop α on X based at x_0, we can define a bijection $G \longrightarrow \pi_1(X)$ mapping $g \mapsto \alpha$.

The *Uniformisation Theorem* can be viewed as a converse to Example 7.16. Recall that a hyperbolic surface S is complete if the metric $\bar{d}$ described above is metrically complete.

Theorem 7.17. Uniformisation theorem *Let X be a complete hyperbolic surface. Then $X = \mathbb{H}/G$ for some torsion free Fuchsian group G.*

As we shall see below, the proof is a consequence of Poincaré's theorem.

Exercise 7.18. *Prove the claim above that if G is a Fuchsian group acting freely on $\mathbb{H}$, then $p : \mathbb{H} \longrightarrow \mathbb{H}/G$ is a covering map. (If needed, see for example [15], p.155).*

7.4.3 The developing map

The proof of the Uniformisation theorem and of the Poincaré theorem are almost the same. To prove them we need the important idea of the **developing map**. For more detail on this section, see [14].

Suppose we are given a hyperbolic structure on a surface S. Let $\widetilde{S}$ be the universal cover of S, which we view as the set of all the homotopy classes of paths on S with fixed base point $x_0 \in S$. The developing map is a map $D : \widetilde{S} \longrightarrow \mathbb{H}$ defined as follows.

Let $\alpha \in \widetilde{S}$ be a path on S with initial point x_0. We can cover α by a finite number of charts (U_i, Φ_i), so that α passes in turn through $U_0, U_1, \ldots, U_n$ say. From the definition of a hyperbolic structure, if $U_i \cap U_j \neq \varnothing$, then there exists $\gamma_{ij} \in \mathrm{Isom}(\mathbb{H})$ such that $\Phi_i(\alpha \cap U_i \cap U_j) = \gamma_{ij}\Phi_j(\alpha \cap U_i \cap U_j)$.

The map D will send α, and simultaneously the open sets U_i, to $\mathbb{H}$ by a process of "analytic continuation". Thus $D_{|(\alpha \cap U_0)}$ is just the map Φ_0.

Since $U_0 \cap U_1 \neq \varnothing$, then there exists $\gamma_{01} \in \mathrm{Isom}(\mathbb{H})$ such that $\Phi_0(\alpha \cap U_0 \cap U_1) = \gamma_{01}\Phi_1(\alpha \cap U_i \cap U_j)$. We set $D_{|(\alpha \cap U_1)} = \gamma_{01}\Phi_1$. This agrees with the previous definition on $U_0 \cap U_1$ and extends the definition of D to $U_0 \cup U_1$.

Continuing in this way, we set $D_{|(\alpha\cap U_2)} = \gamma_{01}\gamma_{12}\Phi_2$, and so on. In this way we obtain a path $D\alpha$ in $\mathbb{H}$. We define the map D mapping $\widetilde{S} \ni \alpha \mapsto D\alpha(1) \in \mathbb{H}$. It is tedious but not hard to check that D is independent of the choice of sequence of charts U_i, and also depends only on the homotopy class of α.

Exercise 7.19. *Check that* $D : \widetilde{S} \longrightarrow \mathbb{H}$ *is a covering map.*

The developing map also induces a map $D_* : \pi_1(S) \longrightarrow \mathrm{Isom}(\mathbb{H})$. Let $\alpha \in \pi_1(S)$ be a loop based at $x_0 \in S$. This means that the final point of α is also x_0, so that in the above construction we may take $U_n = U_0$. Thus $D(\alpha)$ is of the form $g\Phi_0(x_0)$ for some $\gamma = \gamma_{01}\gamma_{12}\dots\gamma_{n-1,n} \in \mathrm{Isom}\,\mathbb{H}$. We define $D_*(\alpha) = \gamma = \gamma_{01}\gamma_{12}\dots\gamma_{n-1,n}$.

Exercise 7.20. *Check the following points:*
(*i*) $D_*(\alpha)$ *does not depend on the choice of* α *in its homotopy class.*
(*ii*) D_* *is a group homomorphism* $\pi_1(S) \longrightarrow \mathrm{Isom}(\mathbb{H})$.

Lemma 7.21. D_* *is an isomorphism onto its image.*

Proof. If $D_*(\alpha) = \mathrm{id}$ then the developing image of α in $\mathbb{H}$ is a loop starting and ending at $\Phi_0(x_0)$. Since $\mathbb{H}$ is simply connected, this loop is null-homotopic. Therefore it lifts (since D is a covering map) to a null homotopy of α in S. □

Exercise 7.22. *Here are some further points to check:*
(*iii*) $G = D_*(\pi_1(S))$ *acts without fixed points in* $\mathbb{H}$.
(*iv*) $G = D_*(\pi_1(S))$ *is a discrete subgroup in* $\mathrm{Isom}(\mathbb{H})$.

7.5 Proof of the Uniformisation Theorem

The proof below is an elaboration of that in [14], see also [11].

If S is a hyperbolic surface we denote its universal cover by $\tilde{S}$ with projection $p : \tilde{S} \longrightarrow S$. The metric on S given locally by the hyperbolic metric in charts. Piecing charts together enables us to define the length $l(\gamma)$ of a path γ. Then we can define a metric on S by

$$d(p,q) = inf_\gamma\{l(\gamma) : \gamma \text{ is a path from } p \text{ to } q\}.$$

It is easy to check that this is the same as the metric $\bar{d}$ described above. We can lift charts and hence the hyperbolic structure on S to a structure on $\tilde{S}$ and then define a metric on $\tilde{S}$ in a similar way. Throughout the proof, we will repeatedly use the path lifting and homotopy lifting properties of covering maps.

The main result we need is the following, which is essentially a version of the Hopf-Rinow theorem, see [11].

Theorem 7.23. *Let S be a hyperbolic surface. The following are equivalent:*

1. *The developing map $D : \tilde{S} \longrightarrow \mathbb{H}$ is a surjective covering map.*
2. *S is metrically complete*
3. *$\tilde{S}$ is metrically complete.*

The **Uniformisation theorem** (Theorem 7.17) is a corollary of this result:

Corollary 7.24. *Every complete hyperbolic surface is $\mathbb{H}/G$ for some torsion free Fuchsian group G.*

Proof. Let $\psi : \pi_1(S) \longrightarrow \operatorname{Isom}\mathbb{H}$ be the holonomy homomorphism and let $G = \psi(\pi_1(S))$. Check that the homotopy lifting property implies that ψ is an isomorphism. (Note that G being torsion free is just the condition for the hyperbolic structure on S to be without cone points; the same proof works allowing cone points if desired.)

We will show that $D : \tilde{S} \longrightarrow \mathbb{H}$ induces an isometry $\bar{D} : S \longrightarrow \mathbb{H}/G$. Since we know that D is a covering map, it will follow (cf Example sheet 6 Q. 6) that G is discrete.

We define $\bar{D}$ as follows. Let $x_0 \in S$ be a basepoint and suppose that $x \in S$. Join x_0 to x by a path α and define $E_\alpha(x) \in \mathbb{H}$ to be the endpoint of the developing image of α. We want to define $\bar{D}(x)$ to be the image of $E_\alpha(x)$ in $\mathbb{H}/G$, so we need to show that the image of $E_\alpha(x)$ in $\mathbb{H}/G$ is independent of the choice of α.

Suppose that β is another such path and let (U, ϕ) be the chart containing x on S. Then $E_\alpha(x) \in g_\alpha\phi(U)$ and $E_\beta(x) \in g_\beta\phi(U)$ for some $g_\alpha, g_\beta \in \operatorname{Isom}\mathbb{H}$. We need to show that $E_\alpha(x) = hE_\beta(x)$ for some $h \in G$. Now $\beta^{-1}\alpha$ is a loop in S based at x_0. We find the developing image of $\beta^{-1}\alpha$ by following the developing image of α to $g_\alpha\phi(U)$, then by following the path obtained by applying $g_\beta^{-1} g_\alpha$ to the developed path from $g_\beta \phi(U)$ back along β to $D(x_0)$. The first chart along this second leg of the path will be $g_\alpha \phi(U)$ and the final chart we get to will be $g_\beta^{-1} g_\alpha \phi(U)$ Since $\beta^{-1}\alpha$ is a loop in S based at x_0, it follows that $g_\beta^{-1} g_\alpha \in G$. Hence $E_\alpha(x), E_\beta(x)$ descend to the same point in $\mathbb{H}/G$.

To see that $\bar{D}$ is injective, we first note that $\bar{D}$ is a covering map. It follows that the number of points in $\bar{D}^{-1}(v)$ is locally constant for $v \in \mathbb{H}/G$. It also follows from the definition that if v_0 is the image of $D(x_0)$, then $D^{-1}(v_0) \subset \mathbb{H}$ is exactly the G-orbit of $D(x_0)$, and hence that $\bar{D}^{-1}(v_0)$ contains exactly one point. The result follows. □

We prove Theorem 7.23 with the following three lemmas. Throughout the proof, it is a good idea to keep in mind is the incomplete structures discussed in Section7.3, which illustrate what might go wrong.

Lemma 7.25. (1) ⇒ (2)

Proof. We first show that (1) implies that closed bounded sets in S are compact.

Since $\tilde{S}$ and $\mathbb{H}$ are simply connected, (1) implies that D is a homeomorphism. By the definition of charts, D is a local isometry and hence an isometry. Since closed bounded sets in $\mathbb{H}$ are compact, the same is true of closed bounded sets in $\tilde{S}$. In particular, any closed ball $\overline{B_r(\tilde{x})}$ for $\tilde{x} \in \tilde{S}$ and $r > 0$ is compact.

Consider a ball $B_r(x) \subset S$ for any $r > 0$ and $x \in S$. Let $\tilde{x} \in p^{-1}(x)$. Suppose that $y \in B_r(x)$. Then there is a path γ from x to y of length $a < r$. This lifts to a path $\tilde{\gamma}$ in $\tilde{S}$ from $\tilde{x}$ to point $\tilde{y}$ with $p(\tilde{y}) = y$ of length a. So $\tilde{y} \in B_r(\tilde{x}) \subset \tilde{S}$. Hence

$$y \in p(B_r(\tilde{x})) \subset p(\overline{B_r(\tilde{x})}).$$

Since $K = p(\overline{B_r(\tilde{x})})$ is the image of the compact set $\overline{B_r(\tilde{x})}$, it is compact. This proves that $B_r(x)$ is contained in the compact set K. Since compact sets are closed, we get that the closure $\overline{B_r(x)} \subset K$ is a closed subset of a compact set which is compact.

Now we can easily prove that (1) ⇒ (2). Suppose that (x_n) is Cauchy sequence in S. Then there exists $r > 0$ such that $x_n \in \overline{B_r(x_0)}$ for all $n \in \mathbb{N}$. Since this set is compact, it is sequentially compact and so (x_n) has a subsequence converging to a limit in S. Since (x_n) is Cauchy, it converges in S. □

Lemma 7.26. (2) $\Rightarrow$ (3)

Proof. Suppose that $(x_n) \subset \tilde{S}$ is Cauchy. Notice that $p : \tilde{S} \longrightarrow S$ decreases distance, in other words $d_S(p(z), p(z')) \leq d_{\tilde{S}}(z, z')$ for all $z, z' \in \tilde{S}$. This is because a path from z to z' in $\tilde{S}$ of length r projects to a path of the same length from $p(z)$ to $p(z')$. Hence $(p(x_n)) \subset S$ is Cauchy in S, hence by our hypothesis (2) has a limit $w \in S$.

Let $B_r(w)$ be an open ball evenly covered by p. (This exists since p is a covering map.) This means that $p^{-1}(B_r(w))$ is a union of disjoint balls V_i each homeomorphic to $B_r(w)$; furthermore by the definition of the hyperbolic structure on $\tilde{S}$ each ball is isometric to $B_r(w)$. The centres $\{p^{-1}(w)\}$ of these balls are separated by a uniform distance at least $2r$. Since (x_n) is Cauchy, all points eventually lie in just one V_i and hence converge to the unique point $\{p^{-1}(w)\} \cap V_i$. $\square$

Lemma 7.27. (3) $\Rightarrow$ (1)

Proof. We will show that D is a surjective local homeomorphism with the path lifting property, and then apply Proposition 7.28 below. That D is a local homeomorphism (every point $x \in \tilde{S}$ is contained in a neighbourhood U such that the restriction of D to U is a homeomorphism onto its image) is immediate from the definition of D.

Suppose that $\alpha : [0, 1] \longrightarrow \mathbb{H}$ is a path. We have to show we can lift α to a path $\tilde{\alpha} : [0, 1] \longrightarrow \tilde{S}$ such that $D \circ \tilde{\alpha} = \alpha$.
(i) Say $t_0 \in [0, 1)$ is such that we can lift $\alpha : [0, t_0]$ to $\tilde{\alpha} : [0, t_0] \longrightarrow \tilde{S}$. It is easy to see that we can extend $\tilde{\alpha}$ to a lift of α on an interval $[0, t_0 + \epsilon)$ for some $\epsilon > 0$. In fact $\tilde{\alpha}(t_0)$ has a neighbourhood U on which D is an isometry. We can choose $\epsilon > 0$ so that $(\alpha(t_0 - \epsilon), \alpha(t_0 + \epsilon)) \subset D(U)$. Then $D^{-1} \circ \alpha$ is well defined on $(t_0 - \epsilon, t_0 + \epsilon)$ and $D^{-1} \circ \alpha$ extends $\tilde{\alpha}$ to $[0, t_0 + \epsilon)$.
(ii) Say $t_0 \in [0, 1)$ is such that we can lift $\alpha : [0, t_0)$ to $\tilde{\alpha} : [0, t_0) \longrightarrow \tilde{S}$. Pick $t_n \longrightarrow t_0$. Then $(\alpha(t_n))$ is Cauchy in S. Since D is a local isometry, $(\tilde{\alpha}(t_n))$ is Cauchy in $\tilde{S}$, hence by (3) has a limit $w \in \tilde{S}$ say. Clearly $D(w) = \alpha(t_0)$ and so defining $\tilde{\alpha}(t_0) = w$ extends $\tilde{\alpha}$ to a path defined on $[0, t_0]$.

Now (i) and (ii) together with a standard argument using connectivity of $[0, 1]$ show that α lifts to a path $\tilde{\alpha} : [0, 1] \longrightarrow \mathbb{H}$.

This implies that D is surjective: suppose that $y \in \mathbb{H}$. Pick $x \in ImD$ and joining x to y by a path α in $\mathbb{H}$ with $\alpha(0) = x, \alpha(1) = y$. Lift α to $\tilde{\alpha}$ and note that $D(\tilde{\alpha}(1)) = y$.

Finally, apply Proposition 7.28 to prove (1). $\square$

Proposition 7.28 is a general result about covering maps:

Proposition 7.28. *Suppose that $f : X \longrightarrow Y$ is a local homeomophism with the path lifting property. Then f is a covering map.*

Proof. This result is standard but since I don't know a straighforward reference, here is a sketch proof.

Let $\tilde{Y}$ be the universal cover of Y with projection map $p : \tilde{Y} \longrightarrow Y$. Let $x_0 \in X$, $y_0 \in Y$ $z_0 \in \tilde{Y}$ be base points with $p(z_0) = f(x_0) = y_0$.

We first claim that p lifts to a map $P : \tilde{Y} \longrightarrow X$ such that $f \circ P = p$. For let $z \in \tilde{Y}$ and let α be a path from z_0 to z. Then $p \circ \alpha$ is a path from y_0 to $p(z)$. By the hypothesis on f, this lifts to a path from x_0 to w say in X. Define $P(z) = w$. We have to check that this is independent of the choice of α. Since $\tilde{Y}$ is simply connected it will be enough if we show it is independent of 'small' homotopies of α, which can be done since f is a local homeomorphism.

Now suppose that $y \in Y$. We have to show that y has a neighbourhood which is evenly covered by f in X. Choose a neighbourhood U of y whose inverse image under p is evenly covered by sets $U_i \subset \tilde{Y}$. We claim that the sets $P(U_i)$ evenly cover U for the map f. It will be enough to show that two sets $P(U_i), P(U_j)$ are either disjoint or coincide.

Suppose that $x \in P(U_i) \cap P(U_j)$. We may assume that U is simply connected. Any point $z \in U$ is connected to $f(x)$ by a path which has a lift β in X; it follows since U is simply connected that the endpoint of β is uniquely defined and also (by lifting β to each of U_i and U_j) lies in both $P(U_i)$ and $P(U_j)$. Thus $P(U_i) = P(U_j)$. □

7.6 Proof of Poincaré's theorem

In this section we outline of how Poincaré's Theorem can be proved in a similar way to Theorem 7.23. Another version of the proof can be found in [10]. First, we recall the statement.

Suppose $\bar{R} \subset \mathbb{H}$ is a closed convex polygon whose sides are identified in pairs by isometries $\{g_1, \ldots, g_r\} = G_0$. Recall the **cycle conditions** from Chapter 6:

(a) the angle sum corresponding to each vertex cycle with the vertex $v \in \mathbb{H}$ is $\sum_{i=0}^{k-1} \theta_i = \frac{2\pi}{p}$ with $p \in \mathbb{N}$ and $(e_{k-1} \cdots e_0)^p = \mathrm{id}$.

(b) if $\bar{R}$ has an ideal vertex $v \in \partial\mathbb{H}$, then the vertex cycle corresponding to v is parabolic.

Theorem 7.29. *(Poincaré's theorem)*
Say $\bar{R}$ is finite sided convex polygon whose sides are identified in pairs by isometries G_0 as above, and suppose the cycle conditions hold. Then:
(i) G_0 generates a discrete group $G \subset \operatorname{Isom}\mathbb{H}$;
(ii) R is a locally finite fundamental domain for the G-action;
(iii) $G =< g_1, \ldots, g_r : \text{vertex relations} >$.

Notice that (ii) shows that the images of R tessellate $\mathbb{H}$. The content of (iii) is that the vertex relations are all one needs for a presentation of G.

To prove Theorem 7.29, we introduce an abstract group G^*, generated by G together with the cycle relations, and an abstract space X, which can think of as a patchwork quilt, in which the patches are copies of $\bar{R}$, one for each $g \in G^*$, sewn together as dictated by the side pairings. We shall show that there is a natural equivariant homeomorphism of X onto $\mathbb{H}$ and that G is isomorphic to G^*.

More precisely, let

$$G^* =< g_1, \ldots, g_r : \text{vertex relations} > .$$

We have a natural group homomorphism $j : G^* \longrightarrow G$, where $G \subset \operatorname{Isom}\mathbb{H}$ is the image of G^* in the isometry group of $\mathbb{H}$. Statement (iii) of Theorem 7.29 is that the map j is an isomorphism.

Also let $\bigsqcup_{g \in G^*}(\bar{R}, g)$ be the disjoint union of copies of $\bar{R}$, one for each element in G^*, and let

$$X = \big(\bigsqcup_{g \in G^*}(\bar{R}, g)\big)/ \approx$$

where $\approx$ is the equivalence relation which identifies points in X according to side pairings and cycles. Thus we identify $x \in (\bar{R}, \mathrm{id})$ with $y \in (\bar{R}, g)$ if $g \in G_0$ and $y = g^{-1}x$, and more generally $x \in (\bar{R}, h)$ is identified to $y \in (\bar{R}, hg)$ if $g \in G_0$ and $y = g^{-1}x$. The relation $\approx$ is the equivalence relation generated by this pairing. (What this means is that we form a patchwork quilt, in which the patches are copies of $\bar{R}$, one for each $g \in G^*$. Two patches $(\bar{R}, g), (\bar{R}, h)$are sewn together along an edge exactly when $h^{-1}g \in G_0$, in other words, when the two patches differ by a side pairing which carries one of the two regions to the other.)

Notice that G^* acts on X by $g \circ (x, h) \longrightarrow (x, gh)$. One has to check that this action respects $\approx$. Since G^* is generated by G_0, it follows that X is connected. We define a metric on X by taking the hyperbolic metric on each copy of $\bar{R}$ and setting the distance between two points to be the length of the shortest path joining them. (Cf the definition of the metric on $R/\sim$ in Section 7.1.) There is a natural map $\Psi : X \longrightarrow \mathbb{H}$ which sends $x \in (\bar{R}, g)$ to $j(g)(x) \in \mathbb{H}$. We check that Ψ is well defined and a covering map onto its image. Clearly, $\Psi(g \circ (x, h)) = j(g)\Psi(x, h)$.

The main point in proving Theorem 7.29 is the following variant of Theorem 7.23. The proof, which we omit here, is similar to that of Theorem 7.23. In fact, the space X is very closely related to the Cayley complex of the group G, as discussed for example in [20] Section 1.3. The Cayley complex is the dual of X. Following [20], see also the discussion about loops in X below, it is not hard to show that X is simply connected and thus can be viewed as the universal cover of S.

Recall from Section 7.1 that $\bar{R}/\sim$ is the hyperbolic surface obtained by gluing $\bar{R}$ using the side pairings.

Theorem 7.30. *With the above set up, the following are equivalent:*

1. *The map* $\Psi : X \longrightarrow \mathbb{H}$ *is a surjective covering map.*
2. $\bar{R}/\sim$ *is metrically complete.*
3. X *is metrically complete.*

Lemma 7.31. *Let* $\bar{R}$ *be a (closed) finite sided polygon with side pairings. Then* $\bar{R}/\sim$ *is metrically complete iff the condition the cycles corresponding to the ideal vertices (cusps) of* $\bar{R}$ *are parabolic.*

Proof. This we leave as an exercise (see discussion in class and also [14]). □

Corollary 7.32. *Suppose that all the cycle conditions hold; in particular that the cycles corresponding to the ideal vertices (cusps) of* R *are parabolic. Then* X *is simply connected and* Ψ *is a bijection onto* $\mathbb{H}$.

Proof. That Ψ is surjective follows from Theorem 7.30 and Lemma 7.31. Then X is simply connected since $\mathbb{H}$ is simply connected and Ψ is a covering map. To show that Ψ is injective: suppose that $\Psi(x) = \Psi(y)$. We can join x to y by a path β in X; then $\Psi \circ \beta$ is a loop in $\mathbb{H}$ which is null homotopic since $\mathbb{H}$ is simply connected. Since Ψ is a covering map, this lifts to a loop in X, so that $x = y$. □

We also have to show that the map $j : G^* \longrightarrow G$ is an isomorphism. Notice that any path in X passes in order through a sequence of regions $(\bar{R}, g_i)$ where $g_i^{-1}g_{i+1} \in G_0$. Thus a path α in X corresponds to an expression $e(\alpha) = e_0e_1 \dots e_{n-1}, e_i \in G_0$, while a path from $(\bar{R}, \mathrm{id})$ to $(\bar{R}, g)$ gives an expression $g = e_0e_1 \dots e_{n-1}, e_i \in G_0$. Similarly, any product of generators corresponds to a path. In particular, if α is a loop in X we find a relation $e(\alpha) = e_0e_1 \dots e_n = \mathrm{id}$. Notice also that a small loop round a vertex of X corresponds to one of the cycle relations in G^*. If a segment of a path α cuts a side of $\bar{R}$ and immediately reenters across the same side, then $e(\alpha) = \dots ee^{-1} \dots$. Thus ensuring (as we may) that this never happens, is equivalent to requiring that the sequence of generators $e(\alpha) = e_0e_1 \dots e_n, e_i \in G_0$ is *reduced*, ie never contains consecutive terms of the form $xx^{-1}, x \in G_0$.

Lemma 7.33. *Suppose that α is a null homotopic loop in X. Then the corresponding path relation $e(\alpha) = e_0e_1 \dots e_n = \mathrm{id}$ is a relation in G^*.*

Proof. Let $\alpha_s, s \in [0, 1]$ be a family of paths such that $\alpha_0 = \alpha$ and $\alpha_1 = \{\mathrm{pt}\}$. Without loss of generality, we may assume all the paths a_s are reduced. As α_s moves across the homotopy, the only times the product $e(\alpha_s)$ changes is when α_s crosses a vertex in X. This changes $e(\alpha_s)$ by a relation coming from a vertex cycle corresponding to a vertex v of $\bar{R}$. In this way, one can prove inductively that the relation $e(\alpha) = \mathrm{id}$ may be built up as a product of vertex relations. □

Corollary 7.34. *The map $j : G^* \longrightarrow G$ is an isomorphism.*

Proof. Suppose that $h \in G^*$ and that $j(h) = \mathrm{id} \in \mathrm{Isom}\,\mathbb{H}$. Take a path in X from $x_0 \in (\bar{R}, \mathrm{id})$ to the corresponding point $(x_0, h) \in (\bar{R}, h)$. This projects to a loop in $\mathbb{H}$. This loop is null homotopic since $\mathbb{H}$ is simply connected, so lifts to a null homotopy in X and hence to a relation $h = \mathrm{id}$ in G^*. □

Chapter 8

The limit set

8.1 The limit set

Definition 8.1. *Let G be a Fuchsian group. The* limit set $\Lambda = \Lambda(G)$ *of G is the closure of the non-elliptic fixed points of G in the Euclidean metric on $\mathbb{D} \cup \partial\mathbb{D}$. Clearly $\Lambda \subset \partial\mathbb{D}$.*

Theorem 8.2. *Let G be a Fuchsian group and let Λ be its limit set. Then:*
(i) Λ is closed and G-invariant.
(ii) $|\Lambda| \leqslant 2 \iff G$ is elementary. If $|\Lambda| > 2$ then $|\Lambda| = \infty$.
If G is non-elementary, then:
(iii) Λ is minimal, that is, every G-orbit is dense.
(iv) Λ is uncountable and perfect.
(v) Either Λ is nowhere dense or $\Lambda = \partial\mathbb{D}$.
(vi) For any $x \in \mathbb{D} \cup \partial\mathbb{D}$, Λ is the set of accumulation points of the orbit Gx.

Recall that if X is a metric space, then a subset $A \subset X$ is *nowhere dense* in X if for all non-empty open sets $U \subset X$, there exists an open non-empty set $V \subset U$ such that $V \cap A = \varnothing$. Equivalently, the interior of the closure of A is empty, $Int\ (\bar{A}) = \emptyset$. A set $A \subset X$ is *perfect* if it is closed and if every point in A is an accumulation point of A, that is, there exist distinct points $a_n \in A$ such that $a_n \longrightarrow a$.

We will need a general fact from metric spaces.

Lemma 8.3. *A perfect subset A of a complete metric space X is uncountable.*

Proof. Recall the Baire category theorem, which says that a complete metric space is not a countable union of nowhere dense subsets. Suppose that $A \subset X$ is perfect. Since A is closed, it is itself a complete metric space so we can apply the Baire category theorem to A. Thus it will be sufficient to see that each point $a \in A$ is nowhere dense in A. Since $\{a\}$ is closed, we have to see that $\{a\}$ has empty interior I. Clearly $I \subset \{a\}$ so if I is non-empty, then $I = \{a\}$. Now I is open in A so $I = U \cap A$ for some set U which is open in X. But A is perfect, so if U is open in X and $a \in U$ then $U \cap A$ contains points in A other than a. This shows that $I = \{a\} \neq U \cap A$ which completes the proof. □

Notation: If $g \in G$ is hyperbolic, we write g^+ for its attracting fixed point. For simplicity of notation, we also write g^+ to denote the unique fixed point if g is parabolic.

Proof. (i) Λ is closed by definition. To prove G-invariance, say $\eta \in \Lambda$. By definition we have $\eta = \lim g_n^+$ where $g_n \in G$ are non-elliptic. Then $gg_n^+ \longrightarrow g\eta$ and $g(g_n^+) = (gg_ng^{-1})^+$, which shows that $g(\eta)$ is also a limit of non-elliptic fixed points.

(ii) If G is elementary, then $|\Lambda| \leqslant 2$ follows from our list of elementary groups. If G is non-elementary, then G has infinitely many distinct hyperbolic axes and hence Λ contains infinitely many distinct hyperbolic fixed points.

(iii) Say $E \subset \Lambda$ is G-invariant and closed. We must show that $E = \Lambda$. We will show that every non-elliptic fixed point is in E, from which the result follows. So let $h \in G$ be non-elliptic. We must show that $h^+ \in E$. Pick $\eta \in E$. We have three cases:
(a) If $\eta = h^+$ there is nothing to prove.
(b) Say $\eta \in E$ and $\eta \neq h^-$. Then $h^n(\eta) \longrightarrow h^+$. So $h^+ \in E$.
(c) Suppose $\eta = h^-$. (This case only happens if h is hyperbolic.) Since G is non-elementary, there exists an element $g \in G$ such that $g(\mathrm{Ax}h) \neq \mathrm{Ax}h$ and so $g(\eta) \neq h^-$. By the G-invariance, $g(\eta) \in E$.
Now arguing as in (a) with $g(\eta)$ in place of η we get that $h^n g(\eta) \longrightarrow h^+$. Since E is closed and G-invariant, it follows that $h^+ \in E$. Hence all hyperbolic fixed points are in E and so $\Lambda \subset E$.

(iv) To show that Λ is perfect, suppose that $\eta \in \Lambda$, so by definition $\eta = \lim g_n^+$ for some sequence of hyperbolic elements $g_n \in G$. If all the g_n are distinct then η is an accumulation point as required. If not, then $\eta = g^+$ for some non-elliptic element $g \in G$. Pick $\xi \in \Lambda, \xi \neq \eta^{\pm}$. Then $h^n(\xi) \longrightarrow \eta$ and all the points $h^n(\xi)$ are distinct as required. That Λ is uncountable now follows from Lemma 8.3

(v) This is proved after (vi).

(vi) This property is rather more subtle. We do it first for $x \in \mathbb{D}$. Since G is non-elementary, we know that Gx is infinite. Hence in the Euclidean topology, Gx must have accumulation points. However by discreteness these cannot be in $\mathbb{D}$. Denoting $\mathcal{A}(Gx)$ the set of all the accumulation points of Gx, we have $\mathcal{A}(Gx) \subset \partial\mathbb{D}$. Moreover since $g^n(x) \longrightarrow g^+$ for all $x \in \mathbb{D}$ we have $g^+ \in \mathcal{A}(Gx)$, so $\mathcal{A}(Gx) \supset \Lambda$.

We also claim that $\mathcal{A}(Gx)$ is independent of x. For suppose that $y \in \mathbb{D}$. Then $d_{\mathbb{H}}(gx, gy) = d_{\mathbb{H}}(x, y)$ for all $g \in G$. Thus as $g_n(x) \longrightarrow \xi$ in the Euclidean metric d_E, we have $d_E(g_n x, g_n y) \longrightarrow 0$ and so $g_n(y) \longrightarrow \xi$ also. So $\mathcal{A}(Gx) = \mathcal{A}(Gy)$.

Thus we have only to show that $\mathcal{A}(Gx) \subset \Lambda$ for some choice of $x \in \mathbb{D}$. To prove this we introduce the *convex hull* $\mathcal{C}(\Lambda)$ of Λ in $\mathbb{D} \cup \partial\mathbb{D}$. We say $V \subset \mathbb{D} \cup \partial\mathbb{D}$ is *hyperbolically convex* if $x, y \in V$ implies that the hyperbolic line joining x to y is contained in V. By definition, $\mathcal{C}(\Lambda)$ is the smallest hyperbolically convex set in $\mathbb{D} \cup \partial\mathbb{D}$. It is also called the *Nielsen region* of G.

Clearly $\mathcal{C}(\Lambda)$ is G-invariant. It is also not hard to see that $\mathcal{C}(\Lambda) \cap \partial\mathbb{D} = \Lambda$. Thus choosing $x \in \mathcal{C}(\Lambda)$, we have $\mathcal{A}(Gx) \subset \mathcal{C}(\Lambda)$ which implies that $\mathcal{A}(Gx) \subset \Lambda$ as required.

Now suppose that $\xi \in \mathbb{D}$. Defining $\mathcal{A}(G\xi)$ in the obvious way, we see as before that $\mathcal{A}(G\xi) \supset \Lambda$. To prove the reverse inclusion, we will show that G acts properly discontinuously on $\partial\mathbb{D} \setminus \Lambda$.

We prove this as follows. There is a natural mapping $r : \partial\mathbb{D} \setminus \Lambda \longrightarrow \mathcal{C}(\Lambda)$, often called the *retraction map*, defined as follows. For $\eta \in \partial\mathbb{D} \setminus \Lambda$, consider horocycles based at η of increasing size. There is a first such horocycle which touches $\mathcal{C}(\Lambda)$, and moreover since $\mathcal{C}(\Lambda)$ is convex the horocycle meets $\mathcal{C}(\Lambda)$ in exactly one point which we define to be $r(\eta)$. It is not hard to see that the map r is continuous and G-equivariant. The action of G on $\mathbb{D}$ and properly discontinuous, so given a neighbourhood $U \ni r(\eta)$ with compact closure (in $\mathbb{D}$), only finitely many G-images of U meet U. The inverse image

$W = r^{-1}(U)$ is an open neighbourhood of η in $\partial\mathbb{D} \setminus \Lambda$ which only meets finitely many G-images of W. It follows that η cannot be an accumulation point of any G-orbit, so in particular $\mathcal{A}(G\xi) \subset \Lambda$.

(v) Say $\Lambda \neq \partial\mathbb{D}$. Let $x \in \partial\mathbb{D} - \Lambda$. Since Λ is closed, there is a neighbourhood N_x of x such that $N_x \subset \partial\mathbb{D} - \Lambda$. By G-invariance, $g(N_x) \cap \Lambda = \varnothing \; \forall g \in G$. Now given an open set U with $U \cap \Lambda \neq \varnothing$, we have to find a non-empty open set $V \subset U$ with $V \cap \Lambda = \varnothing$.

Let $\xi \in U \cap \Lambda$. By (vi), the accumulation points of the orbit of x are exactly Λ so we can choose $g_n \in G$ with $g_n(x) \longrightarrow \xi$. Then eventually $g_n(x) \in U$ and so $g_n(N_x) \cap U \neq \varnothing$. Since $g_n(N_x)$ is open and contains $g_n(x) \in \Lambda$, we are done.

□

In somewhat old terminology, a non-elementary Fuchsian group is said to be *of the first kind* if $\Lambda = \partial\mathbb{D}$ and *of the second kind* otherwise.

Corollary 8.4. *If G is of the second kind, then Λ is homeomorphic to the (middle third) Cantor set.*

Proof. It follows easily from the fact that Λ is nowhere dense, that Λ is totally disconnected. For suppose $\xi, \eta \in \Lambda$. Then $\partial\mathbb{D} \setminus \{\xi, \eta\}$ consists of two connected open intervals W_1, W_2. Since Λ is nowhere dense, we can find $a_i \in W_i, a_i \notin \Lambda, i = 1, 2$. The points a_1, a_2 separate $\partial\mathbb{D}$ into two open intervals U_1, U_2 each containing one of ξ, η. Thus the sets $U_i \cap \Lambda$ disconnect Λ. The result follows from the theorem that every perfect totally disconnected subset of $\mathbb{R}$ is homeomorphic to the Cantor set. □