MT5830
Topics in Geometry and Analysis
(a.k.a. Hyperbolic Geometry)

Jonathan M. Fraser

April 14, 2017

About this course

Hyperbolic geometry is a beautiful subject which blends ideas from algebra, analysis and geometry. I first encountered it during my time as an undergraduate (at St Andrews). I took a course taught by Bernd Stratmann in my final semester in the spring of 2009 and that course was my biggest influence whilst writing this course. In fact I still have the notes I made in 2009 (for which I should probably thank my dad for insisting I hang on to them). Additionally, I am indebted to my friend Tony Samuel for providing me with access to some of Bernd's materials directly. Shortly after delivering his course, Bernd moved to The University of Bremen where he continued working on hyperbolic geometry, until his tragic death in 2015 following a period of ill health. You can find an *In Memorium page* here: `http://www.math.uni-bremen.de/bos/`, with links to various materials.

My other main sources in writing these notes were the books:

1. *The Geometry of Discrete Groups* - Alan Beardon

2. *Hyperbolic Geometry* - James W. Anderson

3. *Fuchsian Groups* - Svetlana Katok;

as well as the lecture notes of my friend (and former colleague at the University of Manchester), Charles Walkden. These notes are available online and can be found here: `http://www.maths.manchester.ac.uk/~cwalkden/hyperbolic-geometry/hyperbolic_geometry.pdf`. Although similar ideas are presented in both courses, I would advise sticking to my notes as our notation, scope and terminology will vary from theirs.

Finally, one more source which I use extensively (and I encourage you to do the same) is an amazing modern phenomenon: 'the Internet'. It seems strange to say, but I am often surprised by how hesitant students are to search for things online. In particular, the relevance and utility of Wikipedia as a mathematical reference are underrated. If, for example, you forget what a Cauchy sequence is, then a good option is to look it up on Wikipedia (but don't tell anyone I said that!)

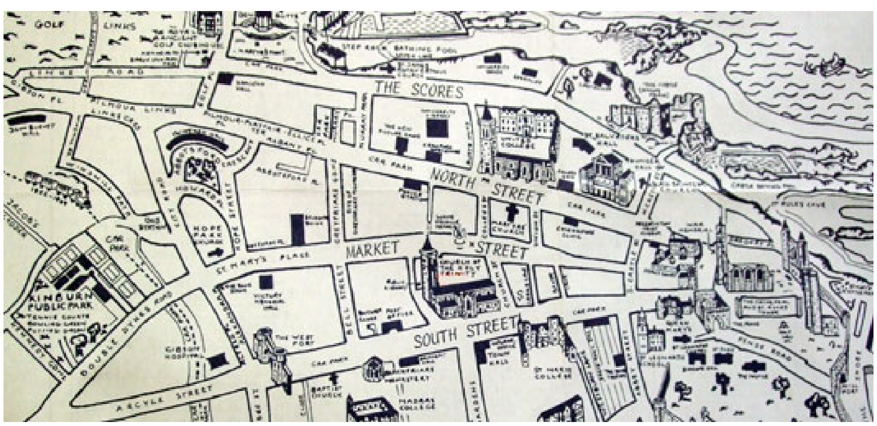
Summer 2009, shortly after taking Hyperbolic Geometry

Now, on to some mathematics... although don't take anything I say in this section too seriously; the course starts in Section 1! In this section I simply want to motivate the topic.

One of the oldest examples of axiomatic mathematics comes from Euclid's postulates on geometry in *The Elements*. These include seemingly harmless things like the ability to join any two points by a straight line. In particular, the *parallel postulate* states:

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

It turns out that this postulate is rather interesting and in some sense defines Euclidean geometry (the geometry of $\mathbb{R}^n$). In two dimensions, there are (in a precise sense) only three possible 'geometries', namely: Euclidean, spherical and hyperbolic (the third being the object of study in this course). Two dimensional spherical geometry refers to the geometry of the surface of a sphere embedded in $\mathbb{R}^3$. Here the 'straight lines' (distance minimising lines) are 'great circles' i.e. circles which lie on the surface of the sphere and have the same diameter as the sphere itself (think of the equator lying on the surface of the earth). Already we see that the *Parallel Postulate* is in trouble because *any* two great circles meet each other in two (antipodal) locations. This can be interpreted as the non-existence of parallel lines in spherical geometry. Parallel lines are essentially unique in Euclidean geometry, in the sense that given a line and a point not on that line, there is a unique line parallel to the given line passing through the given point. This is sometimes known as *Playfair's Axiom* and is equivalent to the *Parallel Postulate*. Hyperbolic geometry is in some sense defined by failing in the opposite direction: given a line and a point not on that line, there there are (continuum) many lines parallel to the given line passing through the given point! In particular, any Euclidean intuition you have concerning parallel lines may have to go out the window!

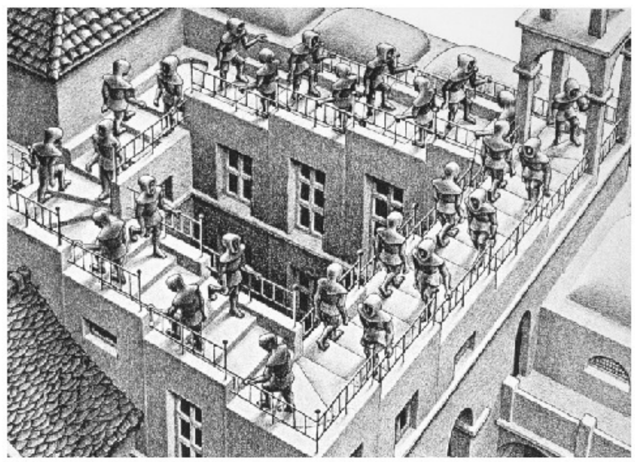


St Andrews: "the town consisting of three parallel streets which all meet at a point"

In higher dimensions, the situation gets rather more complicated. Compared to the three 2-dimensional geometries, it turns out that there are eight 3-dimensional geometries. This was essentially the concern of *Thurston's Geometrisation Conjecture*, proved by Perelman in 2003! Among many other things, the resolution of this problem implied the *Poincaré Conjecture*, which was one of the *Millennium Problems* presented by the Clay Institute in 2000.

Most of the ideas presented in this course, and indeed concerning hyperbolic geometry in general, date back to the 19th Century. Some of the main contributors included (amongst many others): Poincaré, Klein, Möbius, Cayley, Beltrami, etc. As well as being fundamental to the development of geometry, the mathematics developed at this time also had an important influence on physics, philosophy, and even art! Notably, M. C. Escher became renowned for his interpretation of ideas from hyperbolic geometry and general non-Euclidean geometry. In particular, aided by various interactions with (the mathematician) Coxeter, Escher painted many pictures which appeared to defy our usual geometric intuitions. These included things such as the 'never ending stair case' and the 'perpetual motion waterfall'. He also depicted hyperbolic space via a series of beautiful tessellations (see the 'Circle Limit' series).

Circle Limit III, 1959



Ascending and Descending, 1960

Finally, I am grateful to the following team of proof readers (some more mathematical than others): Ailsa Fraser, Iain Fraser, Douglas Howroyd, Nayab Khalid, Rayna Rogowsky and Han Yu. I would also like to thank Tom Elsden and Cara Fraser for not reading the notes.

1 Introduction and some preliminaries

One of the underlying themes of this course is the interplay between the geometry of a metric space and the group of transformations which preserve the metric (group of isometries). Broadly speaking, one may think of 'geometry' as the study of invariants under a fixed group of transformations. In our setting the invariant will be the *hyperbolic metric*, but one may study any metric space in this way, or indeed study other (non-metric) invariants. In this section we will recall and discuss some of the basic concepts we will rely on throughout the course, such as metric spaces, isometries, groups, and group actions.

As with many areas of pure mathematics, we will usually begin with a *space* X, which is just a non-empty set. We are then interested in adding structure to the space (which at the moment is quite boring, at least from a geometers point of view). A *metric* is a function which allows us to measure the distance between two points in our space and the definition simply ensures that the function is a sensible choice.

Definition 1.1. *A* metric space *is a non-empty set X together with a metric $d : X \times X \to \mathbb{R}$ such that*

1. *(positivity)* $\forall x, y \in X : \ d(x, y) \geqslant 0$
2. *(identity of indiscernibles)* $\forall x, y \in X : \ d(x, y) = 0 \Leftrightarrow x = y$
3. *(symmetry)* $\forall x, y \in X : \ d(x, y) = d(y, x)$
4. *(triangle inequality)* $\forall x, y, z \in X : \ d(x, y) \leqslant d(x, z) + d(z, y)$

An important example to keep in mind at this point is Euclidean space $\mathbb{R}^n$ with the usual Euclidean metric $d(x, y) = |x - y|$. In 'reasonable' metric spaces, one can ask the question: 'what is the most efficient way to travel between two points of the space?' In this context 'efficient' means distance minimising in a sense we will make precise. For example, in Euclidean space the most efficient way to travel between two points is via the (unique) straight line joining the two points.

Definition 1.2. *Let (X, d) be a metric space and let $x, y \in X$ with $x \neq y$. A set (curve) $C \subseteq X$ is called a* geodesic *(from x to y) if there exists a continuous bijection $\gamma : [0, 1] \to C$ such that*

1. $\gamma(0) = x$ *and* $\gamma(1) = y$
2. *for all $s, t \in [0, 1]$ we have*

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

Now that we have additional structure on our space we can look for transformations which preserve this structure. Such transformations are called isometries.

Definition 1.3. *Let (X, d) be a metric space. A transformation $\phi : X \to X$ is an* isometry *if*

1. *ϕ is a bijection*
2. *$\forall x, y \in X : \ d(\phi(x), \phi(y)) = d(x, y)$*

Continuing our example from before, the isometries of Euclidean space are precisely maps of the form

$$\phi(x) = t + Ax$$

where $t \in \mathbb{R}^n$ is a translation and A is an orthogonal matrix, i.e. an $n \times n$ matrix over $\mathbb{R}$ such that $AA^T = A^T A = I$.

The collection of all isometries of a metric space is an interesting object in its own right. First of all, note that it is always non-empty and if one composes two isometries, then one gets an isometry back again. This means that the collection of all isometries has an elegant algebraic structure.

Definition 1.4. *A* group *is a set X together with a binary operation which takes a pair $(x, y) \in X \times X$ to another element in X, denoted by $x \cdot y$, which satisfies*

1. *(associativity) $\forall x, y, z \in X : \ x \cdot (y \cdot z) = (x \cdot y) \cdot z$*
2. *(identity) $\exists e \in X : \ \forall x \in X : \ x \cdot e = e \cdot x = x$*
3. *(inverses) $\forall x \in X : \ \exists y \in X : \ x \cdot y = y \cdot x = e$*

The key fact to take away from this section is the following:

Theorem 1.5. *Given any metric space, the collection of all isometries forms a group where the binary operation is composition of maps. This group is sometimes called the group of isometries (of the metric space).*

Composition of maps is very simple However, be careful in which direction you are composing! Given maps $\phi_1, \phi_2 : X \to X$, the composition $\phi_1 \circ \phi_2 : X \to X$ is defined by

$$\phi_1 \circ \phi_2(x) = \phi_1(\phi_2(x))$$

i.e. we are composing from right to left (apply ϕ_2 first, then apply ϕ_1). We say that the group of isometries *acts* on the metric space X and the *action* of an isometry ϕ on an element $x \in X$ is given by $\phi(x)$.

The group of isometries of Euclidean space is called the *Euclidean group* and is given by a semidirect product of $\mathbb{R}^n$ (where the group operation is vector addition) and $O(n, \mathbb{R})$, the group of real $n \times n$ orthogonal matrices. More precisely, the Euclidean group is

$$E(n) \cong O(n, \mathbb{R}) \ltimes \mathbb{R}^n$$

where the semidirect product is the direct product $O(n, \mathbb{R}) \times \mathbb{R}^n$ equipped with the skew-product operation

$$(A_1, t_1) \cdot (A_2, t_2) = (A_1 \circ A_2, A_1(t_2) + t_1).$$

(We will not deal with semidirect products in this course, and so this description is merely an aside). Another important group to mention at the moment is the *general linear group*

$GL(n, \mathbb{R})$, which consists of all real $n \times n$ invertible matrices. Also, the *special linear group* $SL(n, \mathbb{R})$, which consists of all real $n \times n$ matrices with determinant 1. Thus we have the following subgroup hierarchy

$$O(n, \mathbb{R}) \leqslant SL(n, \mathbb{R}) \leqslant GL(n, \mathbb{R}).$$

A common theme in group theory (and related areas) is to consider the following question: once you have a group, can you write down a simple collection of elements which can be used to build all the elements of the group?

Definition 1.6. *Given a group X and a collection of elements $A \subseteq X$, the* subgroup generated by A *is written $\langle A \rangle \leqslant X$ and is defined to be the smallest subgroup of X containing A, i.e.*

$$\langle A \rangle = \bigcap_{A \subseteq Z \leqslant X} Z.$$

A beautiful fact from elementary group theory is that $\langle A \rangle$ is precisely the set of elements which can be written as finite products of elements from A and their inverses. The Euclidean group is generated by reflections in straight lines (infinite extensions of geodesics).

Finally, we end this introduction by recalling the hyperbolic trig functions, which will be relevant for this course. For $x \in \mathbb{R}$ we have

$$\sinh(x) = \frac{e^x - e^{-x}}{2} \qquad \text{and} \qquad \cosh(x) = \frac{e^x + e^{-x}}{2}.$$

Rather than stating various identities concerning these functions, we just emphasise that they all follow from the definitions and you should be ready to derive them when necessary!

2 The Poincaré disk model of hyperbolic space

2.1 The hyperbolic metric

We are about to meet our first model of hyperbolic space and the hyperbolic metric. The space will be something rather familiar: the interior of the unit disk in the complex plane. In particular, we will refer to the Poincaré disk

$$\mathbb{D}^2 = \{z \in \mathbb{C} : |z| < 1\}$$

and the boundary

$$S^1 = \{z \in \mathbb{C} : |z| = 1\} = \overline{\mathbb{D}^2} \setminus \mathbb{D}^2.$$

Of course, the Poincaré disk may be equipped with the Euclidean metric and in this setting it is not difficult to see that its group of isometries is isomorphic to $O(2, \mathbb{R})$ (rotations and reflections). Thus the group of isometries of the disk is a proper subgroup of the group of isometries of the plane due to the fact that there is 'no room' for translations. However, we will put a metric on the disk which recovers the 'room' present in the whole plane. Clearly this metric will have to be very non-Euclidean. The key idea is that as one approaches the boundary S^1 the metric will 'blow-up' in comparison to the Euclidean metric. In other words, moving towards the boundary of the disk will be like moving off to infinity in the plane. For this reason S^1 is sometimes called the *boundary at infinity*.

Consider the following 'hyperbolic kernel' $h : \mathbb{D}^2 \to \mathbb{R}$ defined by

$$h(z) = \frac{2}{1 - |z|^2}.$$

Notice that $h(z)$ blows up as z approaches the boundary. Let $\gamma : [0, 1] \to \mathbb{D}^2$ be a continuous injection given by

$$\gamma(t) = \alpha(t) + i\beta(t)$$

where we assume that α and β are differentiable with continuous derivatives α' and β'. Thus γ defines a curve $C = \gamma([0, 1]) \subset \mathbb{D}^2$ in the Poincaré disk and we may compute the length of the curve with respect to the hyperbolic kernel as

$$L(C) = \int_C h(z)|dz| = \int_C \frac{2\ |d\gamma(t)|}{1 - |\gamma(t)|^2} = \int_0^1 \frac{2\sqrt{\alpha'(t)^2 + \beta'(t)^2}}{1 - (\alpha(t)^2 + \beta(t)^2)}\, dt.$$

This integral looks unpleasant at first sight, but just think of it as the standard 'length' of a smooth curve distorted by the hyperbolic kernel. We are now ready to define the hyperbolic metric. For $w, z \in \mathbb{D}^2$ we say C is a *continuously differentiable* curve joining w, z if C is parameterised by a function γ as above and $\gamma(0) = w$ and $\gamma(1) = z$. The hyperbolic distance between w, z is defined by

$$d_{\mathbb{D}^2}(w, z) = \inf\{L(C) :\ C \text{ is a continuously differentiable curve joining } w, z\}.$$

Again, this looks rather unpleasant at first sight, and not particularly easy to work with! However, as is often the case with complicated abstract objects, we will rarely work directly from this definition. Instead we will derive elementary formulae for the distances between particular points and then explore the metric via its geodesics, isometries, and characteristic features! We will spend the rest of the course trying to understand this beautiful metric and the geometry it creates.

It should be clear that $d_{\mathbb{D}^2}$ satisfies all the conditions of being a metric. Perhaps, the most subtle is 'identity of indiscernibles', but this will be dealt with explicitly later.

2.2 The group of isometries

Our first task will be to identify the isometry group of the metric space $(\mathbb{D}^2, d_{\mathbb{D}^2})$. This turns out to be the collection of all conformal automorphisms of the disk, i.e., angle preserving bijections from the disk to itself. This group has the following simple description:

$$\begin{aligned}\mathrm{con}(1) \quad = \quad \Big\{ g : \mathbb{D}^2 \to \mathbb{D}^2 \ :\ & \text{for some } a, c \in \mathbb{C} \text{ with } |a|^2 - |c|^2 = 1 \\ & g(z) = \frac{az + \bar{c}}{cz + \bar{a}} \ \text{ for all } z \in \mathbb{D}^2 \text{ or} \\ & g(z) = \frac{a\bar{z} + \bar{c}}{c\bar{z} + \bar{a}} \ \text{ for all } z \in \mathbb{D}^2 \Big\}\end{aligned}$$

The distinction between the first and second 'standard form' is simply that the first preserves orientation and the second reverses it. We will write the subgroup consisting of all orientation preserving elements of con(1) as $\mathrm{con}^+(1)$. In fact $\mathrm{con}^+(1)$ is a subgroup of a very important group of transformations acting on the extended complex plane $\hat{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$, known as the *Möbius group* given by

$$\begin{aligned}\text{Möb}^+ \quad = \quad \Big\{ g : \hat{\mathbb{C}} \to \hat{\mathbb{C}} \ :\ & \text{for some } a, b, c, d \in \mathbb{C} \text{ with } ad - bc \neq 0 \\ & g(z) = \frac{az + b}{cz + d} \ \text{ for all } z \in \hat{\mathbb{C}} \Big\}.\end{aligned}$$

Elements of Möb$^+$ are known as *Möbius maps* or *Möbius transformations*. Note that we adopt the convention that, for a Möbius map g, $g(\infty) = a/c \in \mathbb{R}$ if $c \neq 0$ and ∞ otherwise and $g(0) = b/d \in \mathbb{R}$ if $d \neq 0$ and ∞ otherwise. The following fact about Möb$^+$ is a classical result from complex analysis which we will use but not prove.

Theorem 2.1. *Möbius maps are holomorphic bijections from the Riemann sphere (identified with $\hat{\mathbb{C}}$) to itself with non-zero derivative. In particular, they are* conformal*, which means they preserves angles locally.*

The main message from this theorem is that Möbius maps preserve angles. In particular, if two differentiable curves meet at some angle (defined by their tangents at the point of intersection), then the image of these curves under a Möbius map intersect at the same angle.

The following key property will be used throughout the course.

Lemma 2.2. *The image of a (doubly infinite) straight line or a circle under a Möbius map is either a (doubly infinite) straight line or a circle.*

Proof. The proof follows a standard decomposition argument. Given $g \in$ Möb$^+$, we can rewrite $g(z)$ as follows

$$g(z) = \begin{cases} \frac{a}{d}z + \frac{b}{d} & \text{if} \quad c = 0 \\ \frac{bc-ad}{c^2}\frac{1}{z+d/c} + \frac{a}{c} & \text{if} \quad c \neq 0. \end{cases}$$

This shows that if $c = 0$ then g is the composition of a rotation, a dilation and a translation. Also, if $c \neq 0$ then g is the composition of the maps:

- $z \mapsto z + d/c$ (a translation)
- $z \mapsto \frac{1}{z}$ (a circle inversion composed with a reflection)
- $z \mapsto \frac{bc-ad}{c^2} z$ (a rotation combined with a dilation)
- $z \mapsto z + \frac{a}{c}$ (a translation).

Clearly, translations, rotations and dilations send straight lines to straight lines and circles to circles. Perhaps surprisingly this is also almost true for circle inversions. In particular, we will show that the inversion $z \mapsto 1/z$ sends lines and circles to lines and circles. Therefore this is also true of elements in Möb$^+$.

First we will show that circles and straight lines in $\mathbb{C}$ can be expressed in a similar form. First, points $z = x + iy$ on a straight line always satisfy

$$ax + by + c = 0$$

for some $a, b, c \in \mathbb{R}$. This is equivalent to

$$\frac{a}{2}(z + \overline{z}) + \frac{b}{2i}(z - \overline{z}) + c = 0$$

and, setting $B = a/2 + b/(2i)$ and $C = c$, yields

$$Bz + \overline{B}\overline{z} + C = 0.$$

Second, points z on a circle centred at $u \in \mathbb{C}$ with radius r satisfy

$$|z - u|^2 = (z - u)\overline{(z - u)} = r^2.$$

This yields

$$z\overline{z} - z\overline{u} - u\overline{z} + |u|^2 = r^2$$

and, setting $A = 1$, $B = -\overline{u}$ and $C = |u|^2 - r^2$, we get

$$Az\overline{z} + Bz + \overline{B}\overline{z} + C = 0.$$

Thus, this is the general equation for a circle or a straight line, where $B \in \mathbb{C}$, $C \in \mathbb{R}$ and $A \in \{0, 1\}$ with $A = 0$ corresponding to lines and $A = 1$ corresponding to circles.

Therefore if $z \neq 0$ lies on the circle or line corresponding to A, B, C, then dividing by $z\overline{z}$ we obtain

$$C\frac{1}{z\overline{z}} + \overline{B}\frac{1}{z} + B\frac{1}{\overline{z}} + A = 0.$$

If $C = 0$ this shows that $1/z$ lies on the line corresponding to the data "B" $= \overline{B}$, "C" $= A$. If $C \neq 0$, then further dividing by C shows that $1/z$ lies on the circle corresponding to the data "A" $= 1$, "B" $= \overline{B/C}$, "C" $= A/C$. Recall that C is real. We deduce that the image of a circle or a straight line under the inversion $z \mapsto 1/z$ is a circle or a straight line. □

Let us note that some 'obvious' isometries are clearly present in con(1):

1. the identity: choose $a = 1$ and $c = 0$ then

$$g(z) = \frac{1z + \overline{0}}{0 \times z + \overline{1}} = z$$

   is a member of con(1).

2. rotation by an angle θ: choose $a = \exp(i\theta/2)$ and $c = 0$ then

$$g(z) = \frac{\exp(i\theta/2)z + \overline{0}}{0 \times z + \overline{\exp(i\theta/2)}} = \frac{\exp(i\theta/2)z}{\exp(-i\theta/2)} = \exp(i\theta)z$$

is a member of con(1).

3. reflection in the real axis (conjugation): choose $a = 1$ and $c = 0$ then

$$g(z) = \frac{1\overline{z} + \overline{0}}{0 \times \overline{z} + \overline{1}} = \overline{z}$$

is a member of con(1).

4. reflection in any straight line through the origin: this can be built as the composition of rotations and conjugation.

Theorem 2.3. *The elements of* con(1) *form a group. Moreover* con(1) *is generated by* $\mathrm{con}^+(1)$ *together with the conjugation map* $z \mapsto \overline{z}$.

Proof. See the tutorial questions. □

We are almost ready to prove that all elements in con(1) are isometries of $(\mathbb{D}^2, d_{\mathbb{D}^2})$. We begin with a simple technical lemma concerning the derivative of elements in $\mathrm{con}^+(1)$.

Lemma 2.4. *For* $g \in \mathrm{con}^+(1)$ *we have*

$$|g'(z)| = \frac{1 - |g(z)|^2}{1 - |z|^2}.$$

Proof. Let $g \in \mathrm{con}^+(1)$ be given by

$$g(z) = \frac{az + \overline{c}}{cz + \overline{a}}$$

for some $a, c \in \mathbb{C}$ with $|a|^2 - |c|^2 = 1$. Using the quotient-rule, we have

$$|g'(z)| = \left| \frac{a(cz + \overline{a}) - c(az + \overline{c})}{(cz + \overline{a})^2} \right| = \frac{|a|^2 - |c|^2}{|cz + \overline{a}|^2} = \frac{1}{|cz + \overline{a}|^2}.$$

On the other hand we have

$$\begin{aligned}
1 - |g(z)|^2 &= 1 - \left| \frac{az + \overline{c}}{cz + \overline{a}} \right|^2 \\
&= \frac{|cz + \overline{a}|^2 - |az + \overline{c}|^2}{|cz + \overline{a}|^2} \\
&= \frac{(cz + \overline{a})(\overline{cz} + a) - (az + \overline{c})(\overline{az} + c)}{|cz + \overline{a}|^2} \\
&= \frac{(|a|^2 - |c|^2)(1 - |z|^2)}{|cz + \overline{a}|^2} \\
&= \frac{1 - |z|^2}{|cz + \overline{a}|^2}.
\end{aligned}$$

Putting these two estimates together yields the desired result. □

Theorem 2.5. *The elements of* $\mathrm{con}(1)$ *are isometries of* $(\mathbb{D}^2, d_{\mathbb{D}^2})$*, i.e. for* $g \in \mathrm{con}(1)$ *and* $u, v \in \mathbb{D}^2$ *we have*

$$d_{\mathbb{D}^2}(u,v) = d_{\mathbb{D}^2}(g(u), g(v)).$$

Proof. Let $g \in \mathrm{con}^+(1)$ and let C be a continuously differentiable curve joining u and v. Using the substitution $z = g(w)$ and the previous lemma we have

$$\begin{aligned} L(g(C)) = \int_{g(C)} \frac{2}{1-|z|^2}|dz| &= \int_C \frac{2}{1-|g(w)|^2}|g'(w)||dw| \\ &= \int_C \frac{2}{1-|g(w)|^2}\frac{1-|g(w)|^2}{1-|w|^2}|dw| \\ &= \int_C \frac{2}{1-|w|^2}|dw| \\ &= L(C). \end{aligned}$$

It follows from the chain rule that $g \in \mathrm{con}^+(1)$ maps continuously differentiable curves C joining u and v to continuously differentiable curves $g(C)$ joining $g(u)$ and $g(v)$. Therefore

$$d_{\mathbb{D}^2}(u,v) = \inf_C L(C) = \inf_C L(g(C)) = \inf_{g(C)} L(g(C)) = d_{\mathbb{D}^2}(g(u), g(v)).$$

which complete the proof for $g \in \mathrm{con}^+(1)$. However, since elements in $\mathrm{con}(1)$ can be obtained by composition of maps in $\mathrm{con}^+(1)$ and conjugation (which is obviously an isometry), we may conclude the result in full generality. $\square$

The above theorem does not quite say that the isometry group of the Poincaré disk is $\mathrm{con}(1)$. It only says that the isometry group *contains* $\mathrm{con}(1)$. However, it turns out that it is *precisely* the isometry group. We will use this fact, but omit the proof. A proof can be found in Anderson (Proposition 4.2) for example.

2.3 Hyperbolic geodesics

Our next task is to get a better handle on the metric $d_{\mathbb{D}^2}$. We begin by deriving a simple but important formula. Notice how we use the fact that $\mathrm{con}(1)$ is the isometry group.

Lemma 2.6. *For* $z \in \mathbb{D}^2$ *we have*

$$d_{\mathbb{D}^2}(0,z) = \log \frac{1+|z|}{1-|z|}.$$

Proof. Let $z \in \mathbb{D}^2$ and write it in polar form as $z = |z|e^{i\theta}$. We begin by making use of $\mathrm{con}(1)$ to simplify the problem. Let $g \in \mathrm{con}(1)$ be given by

$$g(w) = e^{-i\theta}w.$$

Recall we have already seen that such rotations are members of $\mathrm{con}^+(1)$. Since g is an isometry we have

$$d_{\mathbb{D}^2}(0,z) = d_{\mathbb{D}^2}(g_\theta(0), g_\theta(z)) = d_{\mathbb{D}^2}(0, |z|).$$

Now let C be a continuously differentiable curve joining 0 and $|z|$ parameterised by γ. Then γ can be decomposed into real part and imaginary part, that is $\gamma(t) = \alpha(t) + i\beta(t)$ for real-valued functions $\alpha, \beta : [0,1] \to \mathbb{R}$ such that $\alpha(0) = \beta(0) = \beta(1) = 0$ and $\alpha(1) = |z|$. We then have

$$\begin{aligned} L(C) = \int_0^1 \frac{2\sqrt{\alpha'(t)^2 + \beta'(t)^2}}{1 - (\alpha(t)^2 + \beta(t)^2)}\, dt \quad &\geqslant \quad \int_0^1 \frac{2\sqrt{\alpha'(t)^2}}{1 - \alpha(t)^2}\, dt \\ &= \quad \int_0^1 \frac{2\alpha'(t)\, dt}{1 - \alpha(t)^2} \\ &= \quad \int_0^{|z|} \frac{2\, du}{1 - u^2} \qquad \text{(setting } u = \alpha(t)\text{)} \\ &= \quad \left[\log \frac{1+u}{1-u}\right]_0^{|z|} \\ &= \quad \log \frac{1+|z|}{1-|z|} \end{aligned}$$

proving one direction of the desired result. However, the other direction is simple. There is only one place where we get an inequality rather than an equality and it is easily seen that equality holds here if and only if $\beta(t) = 0 \,\forall t \in [0,1]$. This means that the curve which minimises $L(C)$ must be the Euclidean straight line between 0 and z. Therefore,

$$d_{\mathbb{D}^2}(0, z) = \inf_{\gamma} L(\gamma([0,1])) = \log \frac{1+|z|}{1-|z|}$$

as required. □

We can use the formula from the previous lemma and the con(1) invariance to derive other formulae for the hyperbolic distance between two points.

Corollary 2.7. *The following useful formulae hold for all $z, w \in \mathbb{D}^2$.*

1. $d_{\mathbb{D}^2}(z,w) = \log \frac{|1-z\overline{w}|+|z-w|}{|1-z\overline{w}|-|z-w|}$
2. $\cosh d_{\mathbb{D}^2}(z,w) = \frac{|1-z\overline{w}|^2+|z-w|^2}{|1-z\overline{w}|^2-|z-w|^2}$
3. $\sinh^2 \frac{d_{\mathbb{D}^2}(z,w)}{2} = \frac{|z-w|^2}{(1-|z|^2)(1-|w|^2)}$.

Proof. Let $z, w \in \mathbb{D}^2$ and define $g_w \in \mathrm{con}(1)$ by

$$g_w(u) = \frac{\frac{i}{\sqrt{1-|w|^2}} u + \frac{-iw}{\sqrt{1-|w|^2}}}{\frac{i\overline{w}}{\sqrt{1-|w|^2}} u + \frac{-i}{\sqrt{1-|w|^2}}}.$$

One can check that g_w is indeed in con(1), and that $g_w(w) = 0$. Therefore

$$\begin{aligned} d_{\mathbb{D}^2}(z,w) = d_{\mathbb{D}^2}(g_w(z), g_w(w)) = d_{\mathbb{D}^2}(0, g_w(z)) = \log \frac{1+|g_w(z)|}{1-|g_w(z)|} \quad &= \quad \log \frac{1+|\frac{z-w}{1-z\overline{w}}|}{1-|\frac{z-w}{1-z\overline{w}}|} \\ &= \quad \log \frac{|1-z\overline{w}|+|z-w|}{|1-z\overline{w}|-|z-w|}. \end{aligned}$$

The formulae in 2. and 3. follow immediately by inserting 1. into

$$\cosh d_{\mathbb{D}^2}(z,w) = \frac{e^{d_{\mathbb{D}^2}(z,w)} + e^{-d_{\mathbb{D}^2}(z,w)}}{2} \quad \text{and} \quad \sinh^2 \frac{d_{\mathbb{D}^2}(z,w)}{2} = \frac{e^{d_{\mathbb{D}^2}(z,w)} + e^{-d_{\mathbb{D}^2}(z,w)} - 2}{4}$$

and simplifying. □

Note that this corollary finally establishes that $d_{\mathbb{D}^2}$ is a metric, by showing it satisfies 'identity of indiscernibles'!

The next result is both aesthetically pleasing and fundamental to understanding the geometry of $\mathbb{D}^2$.

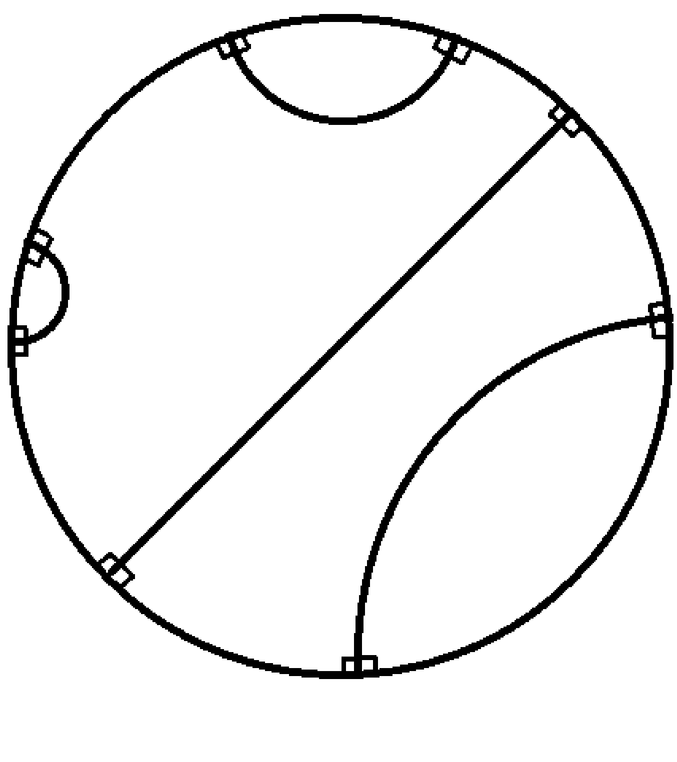
Theorem 2.8. *Hyperbolic geodesics exist and are unique between any two points in $\mathbb{D}^2$. Moreover, they either lie on Euclidean straight lines through the origin or on circles in $\mathbb{D}^2$ orthogonal to S^1 (that is $C \cap \mathbb{D}^2$, where C is a circle orthogonal to S^1).*

Proof. We have already seen that geodesics between 0 and a point $z \in \mathbb{D}^2$ exist and are unique. This extends to any two points $u, v \in \mathbb{D}^2$ by the following simple observation. Consider the element

$$g_u(z) = \frac{\frac{i}{\sqrt{1-|u|^2}}z + \frac{-iu}{\sqrt{1-|u|^2}}}{\frac{i\overline{u}}{\sqrt{1-|u|^2}}z + \frac{-i}{\sqrt{1-|u|^2}}}.$$

Note that g_u in $\mathrm{con}^+(1)$, and note that $g_u(u) = 0$. It is then straightforward to show that the (unique) geodesic between 0 and $g_u(v)$ maps to a geodesic between u and v via $g_u^{-1} \in \mathrm{con}^+(1)$. Also if there were two different geodesics between u and v, then these would map to two different geodesics between 0 and $g_u(v)$ via g_u which would be a contradiction.

We have therefore shown that any geodesic is the image of a straight line emanating from the origin under some $g \in \mathrm{con}(1)$. Therefore, since g is a Möbius transformation, we know that it lies on a circle or a straight line which intersects S^1 at right angles. □

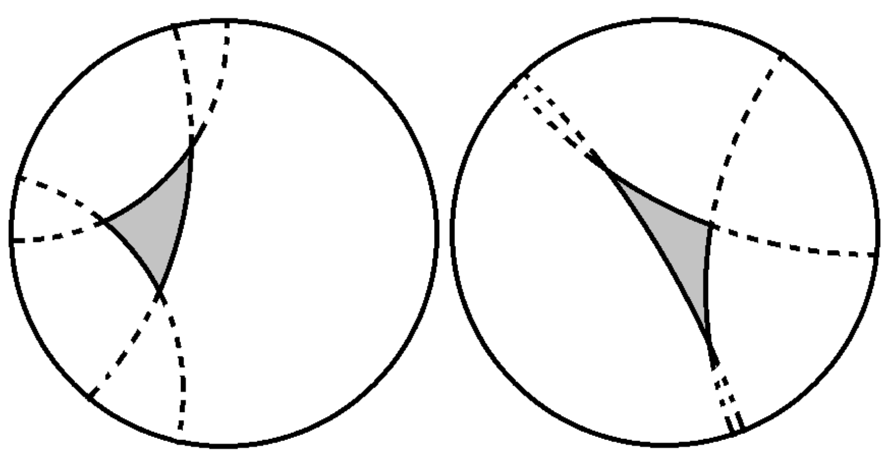


2.4 OK, that's all very nice...

At this point I think it is important to discuss some motivation for this subject. So far we have seen some of the beautiful geometric properties of hyperbolic space. The fact that geodesics lie on circles intersecting the boundary at right angles is particularly pleasing! However, all of this follows from our choice of hyperbolic kernel. In particular, we could have chosen a different kernel: any strictly increasing continuous function $f : [0, 1) \to (0, \infty)$ with the property that $f(x) \to \infty$ as $x \to 1$ would lead to a similar effect if we had chosen the kernel to be $z \mapsto f(|z|)$, although the geometry would not be so beautiful (perhaps?). So what is so special about the hyperbolic kernel and was it just chosen to make pretty pictures? The answer is that the hyperbolic kernel and resulting geometry is *very* special. One can think of kernels which blow up at the boundary as adding 'negative curvature' to the (usually 'flat') complex disk. A good way to visualise this is to imagine the hyperbolic disk as the inside of a cereal bowl. If an ant was to start in the middle and walk to the edge it would find that it needs to make increasingly more effort as it gets nearer to the edge due to the (negative) curvature of the bowl. Whereas walking on the surface of a sphere presents different geometric features due to the (positive) curvature of the sphere. There is a rigorous abstract definition of 'curvature' which can be applied to smooth manifolds and if one wants to study negatively curved spaces then one might ask the following question: what are the symmetric, simply connected, 2-dimensional Riemannian manifolds with constant negative curvature? (The natural question, don't you agree? :-)) The answer is that there is only one, and it is the hyperbolic space we have just constructed! The concepts I have hinted at in this section will not play a role in this course apart from motivating the model. However, the important heuristic to take away is that the hyperbolic kernel was chosen to guarantee *constant* negative curvature. Different kernels would not have achieved this.

2.5 Hyperbolic triangles

Euclidean triangles are some of the first geometric objects we learn about. We learn theorems like Pythagoras' theorem, various results in trigonometry, and formulae for the area. More abstractly, provided your metric space has geodesics, one may consider triangles to be the union of the three geodesics formed by three distinct points in the space. Unsurprisingly, the geometry of hyperbolic triangles (i.e. triangles formed by geodesics in $\mathbb{D}^2$) is rather different from the Euclidean case. However, we already know what they look like, because we understand geodesics.



Here is the hyperbolic analogue of Pythagoras' theorem.

Theorem 2.9. *For every right angled hyperbolic triangle in* $\mathbb{D}^2$ *with hypotenuse of hyperbolic length* c*, and remaining sides of hyperbolic length* a *and* b*, we have*

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

Proof. Since con(1) consists of angle-preserving isometries of $\mathbb{D}^2$, we can assume without loss of generality that 0 is the vertex opposite the hypotenuse, and that the remaining vertices are at α (corresponding to a) and at $i\beta$ (corresponding to b), for $\alpha, \beta \in (0, 1)$. Otherwise we could move the triangle to this special triangle without changing any distances! We have

$$\begin{aligned}
\cosh c = \cosh d_{\mathbb{D}^2}(\alpha, i\beta) &= \frac{|1-\alpha\overline{i\beta}|^2 + |\alpha - i\beta|^2}{|1-\alpha\overline{i\beta}|^2 - |\alpha - i\beta|^2} \\
&= \frac{|1+i\alpha\beta|^2 + |\alpha - i\beta|^2}{|1+i\alpha\beta|^2 - |\alpha - i\beta|^2} \\
&= \frac{1+\alpha^2\beta^2+\alpha^2+\beta^2}{1+\alpha^2\beta^2-\alpha^2-\beta^2} \\
&= \frac{(1+\alpha^2)(1+\beta^2)}{(1-\alpha^2)(1-\beta^2)} \\
&= \cosh d_{\mathbb{D}^2}(0,\alpha)\cosh d_{\mathbb{D}^2}(0,i\beta) \\
&= \cosh a \cosh b
\end{aligned}$$

as required. □

If a, b, c are the lengths of a sides of a triangle (in any metric space), then we always have $c \leqslant a + b$ (this is just the triangle inequality!) Consider a right angled Euclidean isosceles triangle with two sides of length a and the hypotenuse of length c. Then the classical Pythagoras' theorem tells us that $c = \sqrt{2}a$, which tells us that in general the reverse of the triangle inequality can fail by a multiplicative constant, i.e.

$$\frac{1}{\sqrt{2}}(a+a) = c = \sqrt{2}a \leqslant a + a.$$

This is not the case in hyperbolic space and in fact the reverse of the triangle inequality only fails by an *additive* constant for right angled triangles (think about the difference between multiplicative and additive in this setting).

Corollary 2.10. *For every right angled hyperbolic triangle in* $\mathbb{D}^2$ *with hypotenuse of hyperbolic length* c*, and remaining sides of hyperbolic length* a *and* b*, we have*

$$a + b - 2\log 2 \leqslant c \leqslant a + b.$$

Proof. By the triangle inequality we only have to prove the first inequality. Observe that for $x \geqslant 0$ we have

$$\frac{1}{2}e^x \leqslant \cosh x \leqslant e^x,$$

and hence

$$x - \log 2 \leqslant \log\cosh x \leqslant x.$$

Using this general observation, the previous theorem implies

$$\begin{aligned} a+b-2\log 2 = (a-\log 2)+(b-\log 2) &\leqslant \log\cosh a + \log\cosh b \\ &= \log(\cosh a \cosh b) \\ &= \log(\cosh c) \\ &\leqslant \log e^c = c \end{aligned}$$

which is sufficient to prove the theorem. □

An archetypal feature of hyperbolic space is that it has *thin* triangles. Generally speaking, a geodesic metric space is called δ-hyperbolic (not really the same use of the word 'hyperbolic') if it has *thin triangles*, i.e. there exists $\delta > 0$ such that for *any* triangle the distance from at least one of the sides to the opposite vertex is less than δ. Check to see that this fails for Euclidean triangles! Generally, if (X, d) is a metric space and $x \in X$ is a point and $C \subset X$ is a subset (the side of a triangle for example), then the distance from x to C is given by

$$d(x, C) = \inf_{y \in C} d(x, y).$$

Corollary 2.11. *Consider a right angled hyperbolic triangle in $\mathbb{D}^2$. If P is the vertex opposite the hypotenuse, then the hyperbolic distance from P to the hypotenuse is always bounded by $3\log 2$.*

Proof. As in the proof of the previous theorem we can assume without loss of generality that our triangle is of the type considered in that proof. Let L be the straight line from 0 which intersects the hypotenuse of the triangle at right angles. This line naturally splits the hypotenuse into 2 parts which we will label as having lengths c_a (the part joining α) and c_b (the part joining β). In particular, since the sides of a triangle are geodesics we have $c = c_a + c_b$ and moreover the line L splits the original triangle into two smaller right angled triangles. To prove the result it will be sufficient to estimate the length of the line L which we denote by δ. By applying the previous theorem to the original triangle and the two new triangles we obtain:

$$c_a + \delta - 2\log 2 \leqslant a \leqslant c_a + \delta,$$

$$c_b + \delta - 2\log 2 \leqslant b \leqslant c_b + \delta$$

and

$$a + b - 2\log 2 \leqslant c \leqslant a + b.$$

Adding the first to the second and then applying the third we obtain

$$c + 2\delta \leqslant a + b + 4\log 2 \leqslant c + 6\log 2.$$

Canceling c and dividing by 2 completes the proof. If this proof was confusing at all, then draw a picture! □

2.6 Hyperbolic circles

Hyperbolic circles are also very different from Euclidean circles. However, they have a pleasingly straightforward shape: they are Euclidean circles! This similarity and the important differences will become clear throughout this section. We begin by comparing Euclidean and hyperbolic circles centred at 0. For $z \in \mathbb{D}^2$ and $r > 0$, we write

$$C_{\mathbb{D}^2}(z, r) = \{w \in \mathbb{D}^2 : d_{\mathbb{D}^2}(z, w) = r\}$$

for the hyperbolic circle centred at z with radius r and

$$C_E(z, r) = \{w \in \mathbb{D}^2 : |z - w| = r\}$$

for the Euclidean circle centred at z with radius r.

Theorem 2.12. *For any radius $r > 0$ we have*

$$C_{\mathbb{D}^2}(0, r) = C_E(0, \tanh(r/2)).$$

Proof. Suppose $z \in \mathbb{D}^2$ is such that $d_{\mathbb{D}^2}(0, z) = r$. Using the basic formulae we derived above it follows that

$$r = \log \frac{1 + |z|}{1 - |z|}$$

and by rearranging this formula for $|z|$ we obtain

$$|z| = \frac{e^r - 1}{e^r + 1} = \tanh(r/2)$$

as required. □

This result shows us that hyperbolic circles centred at the origin look exactly like Euclidean circles centred at the origin. Since we know the hyperbolic metric is invariant under rotations of $\mathbb{D}^2$ this should not be surprising! We can actually say more simply by using the fact that elements of con(1) send circles to circles (or straight lines).

Theorem 2.13. *Hyperbolic circles in $\mathbb{D}^2$ are Euclidean circles and vice versa (although the centres are different unless at the origin).*

Proof. Let C be a Euclidean circle centred at $z \in \mathbb{D}^2$. Let L denote the (infinite) straight line passing through 0 and the centre of C. This line intersects C at right angles in two points, which we denote by u and v. Moreover, we know that the geodesic between u and v lies on the line L. Let w be the hyperbolic midpoint of this geodesic and let $g \in \text{con}(1)$ be such that $g(w) = 0$. We know $g(C)$ must be a Euclidean circle since g cannot map C to a straight line in this case (think about why this is). Moreover this circle must be centred at the origin since $g(L)$ must be a line passing through $g(u)$, $g(w) = 0$ and $g(v)$ intersecting $g(C)$ at right angles with 0 being the midpoint of the geodesic joining $g(u)$ and $g(v)$. Therefore by the previous theorem it is also a hyperbolic circle. It follows that $C = g^{-1}(g(C))$ is a hyperbolic circle, centred at w.

In the opposite direction, let C be a hyperbolic circle centred at $z \in \mathbb{D}^2$. We do not *a priori* know what C looks like. Let $g \in \text{con}(1)$ be such that $g(z) = 0$. Then $g(C)$ is a hyperbolic circle centred at the origin and is therefore a Euclidean circle. It follows that $C = g^{-1}(g(C))$ is a Euclidean circle. □

It is somewhat of a relief that hyperbolic circles are Euclidean circles as this makes them rather simple to visualise. However, they behave very differently! We will demonstrate this by comparing two important quantities: circumference and area. For $z \in \mathbb{D}^2$ and $r > 0$, we write

$$B_{\mathbb{D}^2}(z,r) = \{w \in \mathbb{D}^2 : d_{\mathbb{D}^2}(w,z) \leqslant r\}$$

and

$$B_E(z,r) = \{w \in \mathbb{D}^2 : |z - w| \leqslant r\}$$

for the hyperbolic and Euclidean ball respectively. Recall that the (Euclidean) circumference of a Euclidean circle is given by the familiar formula

$$L_E(C_E(z,r)) = 2\pi r$$

and the (Euclidean) area of a Euclidean ball is given by the familiar formula

$$A_E(B_E(z,r)) = \pi r^2.$$

The hyperbolic area of a 'reasonable' set F is given by

$$A_{\mathbb{D}^2}(F) = \int_F h(z)^2 |dz| = \int_F \frac{4}{(1-|z|^2)^2} |dz|.$$

Here 'reasonable' refers to a technical integrability condition, which we will not worry about. Certainly 'reasonable' includes any area enclosed by piecewise differentiable curves for example and we will not compute the area of anything that does not fall into this category.

Importantly (and unsurprisingly) hyperbolic area is preserved by con(1).

Theorem 2.14. *Let $F \subset \mathbb{D}^2$ be 'reasonable'. Then for all $g \in \mathrm{con}(1)$ we have*

$$A_{\mathbb{D}^2}(F) = A_{\mathbb{D}^2}(g(F)).$$

Proof. This is similar to the proof that con(1) preserves hyperbolic length and is left as an exercise on the tutorial sheets. □

Theorem 2.15. *For all hyperbolic circles $C_{\mathbb{D}^2}(z,r)$ of radius r, we have*

$$L_{\mathbb{D}^2}(C_{\mathbb{D}^2}(z,r)) = 2\pi \sinh r.$$

Proof. This is in the tutorial sheets. □

As r becomes very large the circumference of a hyperbolic circle grows exponentially in r. More precisely, we have

$$L_{\mathbb{D}^2}(C_{\mathbb{D}^2}(z,r)) = 2\pi \sinh r = 2\pi \frac{e^r - e^{-r}}{2} \sim \pi e^r$$

as $r \to \infty$. Here $A \sim B$ as a parameter $x \to \infty$ formally means that $A/B \to 1$ as $x \to \infty$. This is very different from the Euclidean situation where the circumference grows linearly!

Theorem 2.16. *For all hyperbolic balls $B_{\mathbb{D}^2}(z,r)$ of radius r, we have*

$$A_{\mathbb{D}^2}(B_{\mathbb{D}^2}(z,r)) = 4\pi \sinh^2 \frac{r}{2}.$$

Proof. By using con(1) invariance as usual, we can assume that the ball is centred at 0. If we parameterise $B_{\mathbb{D}^2}(0,r)$ in polar coordinates we get

$$B_{\mathbb{D}^2}(0,r) = \{se^{i\theta} : 0 \leqslant s \leqslant \tanh(r/2) \text{ and } 0 < \theta \leqslant 2\pi\}.$$

Using the substitution $z = se^{i\theta}$, which gives $|dz| = |d(se^{i\theta})| = s\,ds\,d\theta$, we get

$$\begin{aligned} A_{\mathbb{D}^2}(B_{\mathbb{D}^2}(0,r)) = \int_{B_{\mathbb{D}^2}(0,r)} \frac{4}{(1-|z|^2)^2}\ |dz| \ &= \ \int_0^{2\pi}\int_0^{\tanh(r/2)} \frac{4}{(1-s^2)^2} s\,ds\,d\theta \\ &= \ 4\pi \int_0^{\tanh(r/2)} \frac{2s}{(1-s^2)^2}\,ds \\ &= \ -4\pi \int_1^{1-\tanh(r/2)^2} \frac{dt}{t^2} \quad (\text{setting } t = 1-s^2) \\ &= \ 4\pi \frac{\tanh(r/2)^2}{1-\tanh(r/2)^2} \\ &= \ 4\pi \sinh^2 \frac{r}{2} \end{aligned}$$

as required. Note that we used the identity $\cosh(x)^2 - \sinh(x)^2 = 1$ in the final line. □

The area of Euclidean circles grows polynomially (as a square) and again this is very different from the hyperbolic case where we have

$$A_{\mathbb{D}^2}(B_{\mathbb{D}^2}(z,r)) = 4\pi \sinh^2 \frac{r}{2} = 4\pi \frac{(e^{r/2} - e^{-r/2})^2}{4} \sim \pi e^r.$$

A consequence of this is that

$$A_{\mathbb{D}^2}(B_{\mathbb{D}^2}(z,r)) \sim L_{\mathbb{D}^2}(C_{\mathbb{D}^2}(z,r))$$

which is rather strange!

3 The upper half-plane model of hyperbolic space

So far we have been working in the metric space $(\mathbb{D}^2, d_{\mathbb{D}^2})$, which is one model of 2-dimensional hyperbolic space. However, it is not the *only* model of this important space, as we shall see in this section. Understanding and describing different models is important because some problems are more easily understood in one model, over another. This time our ambient space will be the *upper half-plane*

$$\mathbb{H}^2 = \{z \in \mathbb{C} : \mathrm{Im}(z) > 0\} = \{x + iy : y > 0\}$$

and this time the 'boundary at infinity' will be the one point compactification of the real line given by

$$\hat{\mathbb{R}} = \mathbb{R} \cup \{\infty\}.$$

We could now write down an appropriate 'hyperbolic kernel' and define the corresponding hyperbolic metric and hope that this new metric space is compatible with the Poincaré disk model. However, since we already have the Poincaré disk model, it is more straightforward to 'steal' the metric $d_{\mathbb{D}^2}$ via an appropriate Möbius transformation which takes the upper half-plane to the disk. This transformation is known as *the Cayley map* and is defined by

$$\phi(z) = \frac{z - i}{z + i}.$$

One can verify the following basic properties of this map.

Lemma 3.1. *We have*

1. *ϕ is a Möbius map and therefore is conformal (angle preserving) and maps circles and straight lines to circles or straight lines.*
2. *$\phi(\mathbb{H}^2) = \mathbb{D}^2$*
3. *$\phi(\mathbb{R}) = S^1 \setminus \{1\}$ and $\phi(\infty) = 1$*
4. *ϕ^{-1} is also a Möbius map and is given by*

$$\phi^{-1}(z) = -i\,\frac{z + 1}{z - 1}$$

5. *in order to visualise $\mathbb{H}^2$ it is also useful to compute some points precisely, for example: $1 = \phi(\infty)$, $-1 = \phi(0)$, $i = \phi(-1)$, and $-i = \phi(1)$.*

Proof. The non-trivial parts of this lemma are on the tutorial sheet. □

In order to visualise what the Cayley map is doing, observe that we can decompose it into the following three parts defined by

$$T_1(z) = \overline{z} \qquad \text{(reflection in } \mathbb{R}),$$

$$T_2(z) = i + \left(\frac{\sqrt{2}}{|z - i|}\right)^2 (z - i) \qquad \text{(reflection in the circle } C(i, \sqrt{2}))$$

and

$$T_3(z) = -iz \qquad \text{(clockwise rotation by } \pi/2).$$

It is an exercise on the tutorial sheet to show that

$$\phi(z) = T_3(T_2(T_1(z))).$$

We define a metric on $\mathbb{H}^2$ simply by

$$d_{\mathbb{H}^2}(u, v) = d_{\mathbb{D}^2}(\phi(u), \phi(v))$$

for $u, v \in \mathbb{H}^2$. Note that ϕ is automatically an isometry between $(\mathbb{H}^2, d_{\mathbb{H}^2})$ and $(\mathbb{D}^2, d_{\mathbb{D}^2})$. Using the fact that ϕ is both an isometry and a Möbius map we can transfer our knowledge of the metric structure of $\mathbb{D}^2$ to $\mathbb{H}^2$.

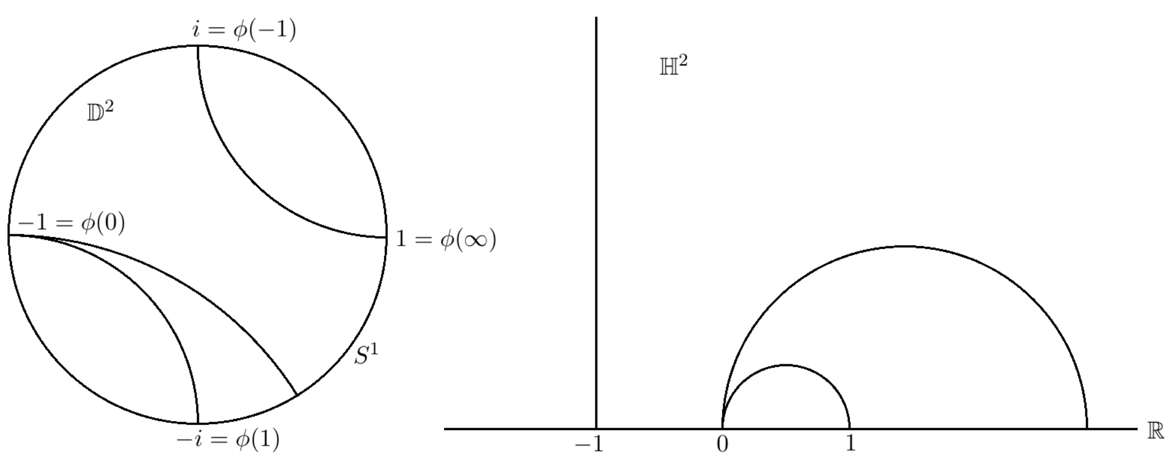
Theorem 3.2. *Hyperbolic geodesics exist and are unique between any two points in $\mathbb{H}^2$. Moreover, they either lie on vertical Euclidean straight lines or on 'half-circles' in $\mathbb{H}^2$ which meet $\mathbb{R}$ at right angles.*

Proof. Let $u, v \in \mathbb{H}^2$. It follows that there is a unique geodesic between u and v and that this geodesic is equal to the image under ϕ^{-1} of the geodesic between $\phi(u)$ and $\phi(v)$ in $\mathbb{D}^2$. This geodesic lies on a circle (or a straight line through the origin) which meets S^1 at right angles in two places. There are now two cases:

1. Neither of the points of intersection of the geodesic with S^1 is equal to 1. In this case the image of the two points of intersection under ϕ^{-1} lie in $\mathbb{R}$ and therefore the image of the circle (or straight line) which the geodesic lies on is a circle which intersects $\mathbb{R}$ at right angles in two places.

2. One of the points of intersection of the geodesic with S^1 is equal to 1. In this case the image of this point under ϕ^{-1} is ∞ and the image of the other point is somewhere in $\mathbb{R}$. Therefore the image of the circle (or straight line) which the geodesic lies on is a vertical Euclidean line which intersects $\mathbb{R}$ at right angles in one place.

The proof is now complete. $\square$



We can also deduce the isometry group of $\mathbb{H}^2$ by using con(1) and ϕ. It turns out to be another important group of Möbius maps: the *projective special linear group* denoted by $\mathrm{PSL}(2, \mathbb{R})$.

Theorem 3.3. *The orientation preserving isometries of* $(\mathbb{H}^2, d_{\mathbb{H}^2})$ *are given by*

$$\begin{aligned} \mathrm{PSL}(2,\mathbb{R}) &:= \Big\{ g : \mathbb{H}^2 \to \mathbb{H}^2 \ : \ \textit{for some } a,b,c,d \in \mathbb{R} \textit{ with } ad-bc = 1 \\ &\qquad\qquad g(z) = \frac{az+b}{cz+d} \ \textit{ for all } z \in \mathbb{H}^2 \Big\} \\ &= \phi^{-1}\mathrm{con}^+(1)\,\phi. \end{aligned}$$

Moreover, the orientation reversing isometries are given by compositions of elements of $\mathrm{PSL}(2,\mathbb{R})$ *with reflection in the imaginary axis:*

$$z \mapsto -\overline{z}.$$

In other words they are maps of the form

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

where $a,b,c,d \in \mathbb{R}$ *with* $ad-bc=1$*.*

Proof. We will deal with the orientation preserving case. The orientation reversing case is in the tutorial sheets. In order to see that elements of $\phi^{-1}\mathrm{con}^+(1)\,\phi$ are isometries of $\left(\mathbb{H}^2, d_{\mathbb{H}^2}\right)$, consider $f = \phi^{-1}g\phi$, for $g \in \mathrm{con}^+(1)$. It follows that for all $z, w \in \mathbb{H}^2$ we have

$$\begin{aligned} d_{\mathbb{H}^2}(z,w) = d_{\mathbb{D}^2}(\phi(z),\phi(w)) &= d_{\mathbb{D}^2}(g(\phi(z)), g(\phi(w))) \qquad (\text{since } g \in \mathrm{con}^+(1)) \\ &= d_{\mathbb{H}^2}(\phi^{-1}(g(\phi(z))), \phi^{-1}(g(\phi(w)))) \\ &= d_{\mathbb{H}^2}(f(z), f(w)) \end{aligned}$$

as required. Moreover, *all* orientation preserving isometries of $\mathbb{H}^2$ are in $\phi^{-1}\mathrm{con}^+(1)\,\phi$ since if g is an orientation preserving isometry of $\mathbb{H}^2$, then a similar argument to the above shows $\phi g\phi^{-1}$ is an orientation preserving isometry of $\mathbb{D}^2$ and therefore in $\mathrm{con}^+(1)$. It follows that $g \in \phi^{-1}\mathrm{con}^+(1)\,\phi$.

We now need to show that $\mathrm{PSL}(2,\mathbb{R}) = \phi^{-1}\mathrm{con}^+(1)\,\phi$. Let $g \in \mathrm{con}^+(1)$ be given by

$$g(z) = \frac{az+\overline{c}}{cz+\overline{a}}$$

for some $a, c \in \mathbb{C}$ with $|a|^2 - |c|^2 = 1$. Then (via some hideous but straightforward rearranging) we obtain

$$\begin{aligned} \phi^{-1}g\phi(z) &= -i\,\frac{\frac{a\left(\frac{z-i}{z+i}\right)+\overline{c}}{c\left(\frac{z-i}{z+i}\right)+\overline{a}} + 1}{\frac{a\left(\frac{z-i}{z+i}\right)+\overline{c}}{c\left(\frac{z-i}{z+i}\right)+\overline{a}} - 1} \\ &= \frac{1}{i} \times \frac{a(z-i)+\overline{c}(z+i)+c(z-i)+\overline{a}(z+i)}{a(z-i)+\overline{c}(z+i)-c(z-i)-\overline{a}(z+i)} \\ &= \frac{z(a+\overline{a}+c+\overline{c})+i(\overline{a}-a+\overline{c}-c)}{zi(a-\overline{a}+\overline{c}-c)+(\overline{a}+a-\overline{c}-c)}. \end{aligned}$$

Note that all four coefficients in this expression are real. It remains to show that they satisfy the required identity:

$$\begin{aligned}
&(a+\overline{a}+c+\overline{c})\times(\overline{a}+a-\overline{c}-c)-(i(\overline{a}-a+\overline{c}-c))\times(i(a-\overline{a}+\overline{c}-c))\\
&\quad= (2\mathrm{Re}(a)+2\mathrm{Re}(c))(2\mathrm{Re}(a)-2\mathrm{Re}(c))+(2\mathrm{Im}(a)+2\mathrm{Im}(c))(2\mathrm{Im}(a)-2\mathrm{Im}(c))\\
&\quad= 4(\mathrm{Re}(a)^2-\mathrm{Re}(c)^2)+4(\mathrm{Im}(a)^2-\mathrm{Im}(c)^2)\\
&\quad= 4(|a|^2-|c|^2)\\
&\quad= 4.
\end{aligned}$$

But wasn't it supposed to be 1? Yes, but we can simply divide top and bottom of the fraction by 2 to obtain a map of the desired form! We have therefore proved that

$$\phi^{-1}\mathrm{con}^+(1)\,\phi \quad\subseteq\quad \mathrm{PSL}(2,\mathbb{R})$$

and it remains to prove the reverse inclusion. This is done similarly, by proving that for any $g\in\mathrm{PSL}(2,\mathbb{R})$ we have $\phi g\phi^{-1}\in\mathrm{con}^+(1)$. We omit the details as they appear in the tutorial sheets. □

It turns out that the hyperbolic kernel needed to defined the metric $d_{\mathbb{H}^2}$ is

$$h_{\mathbb{H}^2}(z)=\frac{1}{\mathrm{Im}(z)}.$$

Theorem 3.4. *For all $u,v\in\mathbb{H}^2$ we have*

$$d_{\mathbb{H}^2}(u,v)=\inf\left\{\int_C\frac{|dz|}{\mathrm{Im}(z)}\ :\ C \text{ is a continuously differentiable curve joining } u,v\right\}.$$

Proof. We have

$$\begin{aligned}
d_{\mathbb{H}^2}(u,v) &= d_{\mathbb{D}^2}(\phi(u),\phi(v))\\
&= \inf\Bigg\{\int_C\frac{2|\phi'(z)||dz|}{1-|\phi(z)|^2}\ :\\
&\qquad\qquad \phi(C)\text{ is a continuously differentiable curve joining }\phi(u),\phi(v)\Bigg\}\\
&= \inf\left\{\int_C\frac{|dz|}{\mathrm{Im}(z)}\ :\ C\text{ is a continuously differentiable curve joining }u,v\right\}
\end{aligned}$$

where the final equality is obtained using the quotient rule and writing $z=x+iy$ as follows

$$\begin{aligned}
\frac{2|\phi'(z)|}{1-|\phi(z)|^2} &= 2\frac{|(z+i)-(z-i)|}{|z+i|^2}\cdot\frac{1}{1-\frac{|z-i|^2}{|z+i|^2}}\\
&= \frac{4}{|z+i|^2-|z-i|^2}\\
&= \frac{4}{x^2+(y+1)^2-(x^2+(y-1)^2)}\\
&= \frac{4}{4y}\\
&= \frac{1}{\mathrm{Im}(z)}
\end{aligned}$$

as required. □

3.1 Hyperbolic triangles revisited

We are now going to prove a fundamental result in hyperbolic geometry concerning hyperbolic triangles. It turns out that this theorem is easier to prove in the upper half-plane than in the Poincaré disk and for this reason we delayed discussion of it until now. This should be taken as a beautiful demonstration of the utility of having two (or more) models of hyperbolic space in our armoury!

The hyperbolic area of a 'reasonable' set F in $\mathbb{H}^2$ is given by

$$A_{\mathbb{H}^2}(F) = \int_F \frac{|dz|}{\text{Im}(z)^2}.$$

Unsurprisingly, this is invariant under the Cayley map, i.e.

$$A_{\mathbb{H}^2}(F) = A_{\mathbb{D}^2}(\phi(F)).$$

This can be found on the tutorial sheets. Therefore we may prove results about triangles in either model by proving them in $\mathbb{H}^2$. Note that triangles can have interior angles equal to 0, but this only happens at vertices on the boundary at infinity! Triangles where all three vertices are on the boundary are called *ideal triangles*. Allowing triangles to have vertices on the boundary is a natural extension to our discussions on triangles (and particularly useful in proving the next theorem)! The 'geodesic' between two points on the boundary is defined to be the unique doubly infinite geodesic ray joining them and the 'geodesic' between a point in $\mathbb{H}^2$ and a point on the boundary is the unique half infinite geodesic ray joining them. By 'geodesic ray' we mean an infinite extension of a geodesic segment to the boundary.

Theorem 3.5 (Gauss-Bonnet Theorem)**.** *For an arbitrary hyperbolic triangle Δ in $\mathbb{H}^2$ with angles $\alpha, \beta, \gamma \geqslant 0$ we have*

$$A_{\mathbb{H}^2}(\Delta) = \pi - (\alpha + \beta + \gamma).$$

Proof. First we prove the result for right angled triangles which have at least one angle equal to zero. Then we will prove it for all triangles which have at least one angle equal to zero, and finally for arbitrary triangles.

Case 1: Let Δ be a triangle with angles $\alpha \geqslant 0$, $\beta = 0$ and $\gamma = \frac{\pi}{2}$.

Without loss of generality we can assume that $i \in \mathbb{H}^2$ is the vertex associated with γ, that the side opposite β is part of the unit circle, and that the remaining two sides are part of 'vertical' geodesics (i.e., the vertex associated with β is ∞). We can achieve this reduction by mapping the vertex associated with γ to i under an element of $\text{PSL}(2, \mathbb{R})$ and then rotating the space around i until the vertex associated with β is at infinity. If you are worried about this rotation, think of the Poincaré disk model where we are simply performing Euclidean rotation around the origin.

The advantage of this reduction is that Δ has a simple parameterisation given by

$$\Delta = \{x + iy \in \mathbb{H}^2 : 0 \leqslant x \leqslant \cos\alpha, \sqrt{1 - x^2} \leqslant y < \infty\}.$$

Using the substitution $x = \cos u$, we get

$$\begin{aligned}
A_{\mathbb{H}^2}(\Delta) = \int_0^{\cos\alpha} \int_{\sqrt{1-x^2}}^{\infty} \frac{1}{y^2}\, dy\, dx = \int_0^{\cos\alpha} \frac{1}{\sqrt{1-x^2}}\, dx \quad &= \quad \int_{\frac{\pi}{2}}^{\alpha} \frac{-\sin u}{\sqrt{1-\cos^2 u}} du \\
&= \quad -\int_{\frac{\pi}{2}}^{\alpha} du \\
&= \quad -[u]_{\pi/2}^{\alpha} \\
&= \quad \frac{\pi}{2} - \alpha \\
&= \quad \pi - \left(\frac{\pi}{2} + \alpha + 0\right)
\end{aligned}$$

as required. Despite this being a very special type of triangle, the hard work is now done as we can build any triangle from triangles of this type.

Case 2: Let Δ be a triangle with $\beta = 0$, and $\alpha, \gamma \geqslant 0$ arbitrary.

Without loss of generality we can assume that the vertex associated with β is at infinity (as before). Let C be the geodesic ray containing the side opposite β (which is necessarily a semi-circle meeting $\mathbb{R}$ at right angles) and let $u \in \mathbb{H}^2$ be the point at which C intersects the vertical line L emanating from the centre of the circle containing C. We now have two sub-cases.
Sub-case (i): Suppose $u \in \Delta$. In this situation the line L splits Δ into two right angled triangles, say Δ_α and Δ_γ where Δ_α contains the vertex associated with α and Δ_γ contains the vertex associated with γ. Since Δ_α and Δ_γ are of the type considered in Case 1, we deduce

$$A_{\mathbb{H}^2}(\Delta_\alpha) = \pi - \left(\frac{\pi}{2} + \alpha\right) \text{ and } A_{\mathbb{H}^2}(\Delta_\gamma) = \pi - \left(\frac{\pi}{2} + \gamma\right).$$

and hence

$$A_{\mathbb{H}^2}(\Delta) = A_{\mathbb{H}^2}(\Delta_\alpha) + A_{\mathbb{H}^2}(\Delta_\gamma) = \pi - \left(\frac{\pi}{2} + \alpha\right) + \pi - \left(\frac{\pi}{2} + \gamma\right) = \pi - (\alpha + \gamma)$$

as required.

Sub-case (ii): Suppose $u \notin \Delta$. We assume without loss of generality that the vertex associated with α lies to the left of the vertex associated with γ and that u lies to the right of both of these points. Therefore the line L lies completely to the right of Δ and we may form a new triangle Δ_0 which has vertices at u, the vertex associated with α, and ∞ (the vertex associated with β. Thus L splits Δ_0 into two triangles, one of which is equal to Δ and the other we denote by Δ_1. In particular, $\Delta_0 = \Delta \cup \Delta_1$ and the triangles Δ_0 and Δ_1 are both of the type considered in Case 1 above (since they both have a right angle at u). We conclude that

$$A_{\mathbb{H}^2}(\Delta_0) = \pi - \left(\frac{\pi}{2} + \alpha\right) = \frac{\pi}{2} - \alpha \text{ and } A_{\mathbb{H}^2}(\Delta_1) = \pi - \left(\frac{\pi}{2} + (\pi - \gamma)\right) = \gamma - \frac{\pi}{2}.$$

and hence

$$A_{\mathbb{H}^2}(\Delta) = A_{\mathbb{H}^2}(\Delta_0) - A_{\mathbb{H}^2}(\Delta_1) = \frac{\pi}{2} - \alpha \ - \ \left(\gamma - \frac{\pi}{2}\right) = \pi - (\alpha + \gamma)$$

as required.

Case 3: Let Δ be a triangle with $\alpha, \beta, \gamma > 0$ arbitrary.

Let a, b, c be the sides of Δ, chosen such that a is opposite to α, b is opposite to β and c is opposite to γ. Extend the geodesic corresponding to the side a from the vertex at γ until it meets the boundary of $\mathbb{H}^2$ at some point, say $\xi \in \mathbb{R} \cup \{\infty\}$. We have formed a new triangle, which we denote by Δ_0, with sides given by: the geodesic ray between the vertex at α and ξ, the extension of a to ξ, and the side of Δ opposite γ (which we have labelled by c).

The side b splits Δ_0 into two triangles, one of which is the original triangle Δ, and we will denote the other by Δ_1. In particular, $\Delta_0 = \Delta \cup \Delta_1$. Clearly, Δ_0 and Δ_1 both have one angle equal to zero (at ξ), and hence are of the type already considered. With θ referring to the one unknown angle of Δ_1, we have

$$\begin{aligned} A_{\mathbb{H}^2}(\Delta) &= A_{\mathbb{H}^2}(\Delta_0) - A_{\mathbb{H}^2}(\Delta_1) \\ &= \pi - (\beta + (\alpha + \theta)) - (\pi - (\theta + (\pi - \gamma))) \\ &= \pi - (\alpha + \beta + \gamma). \end{aligned}$$

which completes the proof. □

The Gauss-Bonnet theorem has some interesting consequences, which reveal some counter intuitive properties of hyperbolic space.

Corollary 3.6. *For any hyperbolic triangle Δ with angles α, β, γ we have*

1. $\alpha + \beta + \gamma < \pi$
2. $A_{\mathbb{H}^2}(\Delta) \leqslant \pi$

Here is another archetypal theorem in hyperbolic geometry, which is orthogonal to Euclidean intuition. It states that angles determine a triangle up to isometry. This is very much false in Euclidean space. Any triangle can be scaled up arbitrarily whilst preserving the angles!

Theorem 3.7. *Let Δ_1 and Δ_2 be hyperbolic triangles which both have interior angles $\alpha, \beta, \gamma \geqslant 0$. Then there exists a hyperbolic isometry g such that $g(\Delta_2) = \Delta_1$.*

Proof. Let Δ_1 be fixed and let z denote the vertex of Δ_1 at the angle α. Let g be the hyperbolic isometry composed of the following maps:

1. an isometry which orients the angles of Δ_2 in the same way as Δ_1. This can be the identity if the orientation is already the same, or an arbitrary orientation reversing isometry if the orientation is different.

2. an isometry which maps the vertex of Δ_2 at α to z

3. the isometry which rotates around z until the sides of Δ_2 emanating from z point in the same directions as the corresponding sides of Δ_1.

We claim that $g(\Delta_2) = \Delta_1$. Suppose these triangles do not coincide and observe that this forces either one of the following situations, both of which lead to a contradiction:

1. one of the triangles is a proper subset of the other. This contradicts the Gauss-Bonnet Theorem, which asserts that both triangles have the same area.

2. the sides of the triangles opposite z are distinct and meet at a point w which is not a vertex of either triangle. Consider the triangle with vertices at w and the vertices of $g(\Delta_2)$ and Δ_1 corresponding to β. The internal angles of this triangle are easily seen to sum to at least π which is a contradiction. Indeed, one of them is β and another is $\pi - \beta$.

The proof is complete. $\square$

4 Classification of hyperbolic isometries

In this section we will take a closer look at the isometries of $\mathbb{H}^2$ (and thus $\mathbb{D}^2$). It turns out that elements in $\mathrm{PSL}(2,\mathbb{R})$ fall into three natural classes: hyperbolic, parabolic and elliptic. These classes display very distinct properties. We will classify and study them by analysing their fixed points, action on hyperbolic space, and standard form. However, the simplest way to determine which class an element of $\mathrm{PSL}(2,\mathbb{R})$ belongs to is by its 'trace'. Let $g \in \text{Möb}^+$ be defined by

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

for some $a, b, c, d \in \mathbb{C}$ with $ad - bc \neq 0$. The *trace* of g is given by

$$\mathrm{tr}(g) = a + d.$$

As stated, the trace is not well-defined because there are different representations of the same Möbius map. However, we adopt the convention that the representation has been normalised such that $ad - bc = 1$ and that (for example) $\arg(a) \in [0, \pi)$. In practice, we will only use the *square* of the trace of elements in $\mathrm{PSL}(2,\mathbb{R})$ or $\mathrm{con}^+(1)$ and so this normalisation is not required, provided the element is written in standard form!

Despite how simple the trace is, it gives a remarkable amount of information about the action of g. We say a non-identity element $g \in \mathrm{PSL}(2,\mathbb{R})$ is:

1. hyperbolic, if $\mathrm{tr}(g)^2 > 4$
2. parabolic, if $\mathrm{tr}(g)^2 = 4$
3. elliptic, if $\mathrm{tr}(g)^2 < 4$.

Theorem 4.1. *Let $g \in \mathrm{PSL}(2,\mathbb{R})$ be any non-identity element. Then*

1. *if g is hyperbolic, then g has precisely two fixed points in $\hat{\mathbb{C}}$, both of which lie in $\mathbb{R} \cup \{\infty\}$*
2. *if g is parabolic, then g has precisely one fixed point in $\hat{\mathbb{C}}$, which lies in $\mathbb{R} \cup \{\infty\}$*
3. *if g is elliptic, then g has precisely two fixed points in $\hat{\mathbb{C}}$, one of which lies in $\mathbb{H}^2$ and the other in $\overline{\mathbb{H}^2} = \{\overline{z} : z \in \mathbb{H}^2\}$.*

Proof. Let $g \in \mathrm{PSL}(2,\mathbb{R})$ be defined by

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

for some $a, b, c, d \in \mathbb{R}$ with $ad - bc = 1$. Suppose g is not the identity and $z \in \hat{\mathbb{C}}$ is a fixed point of g, i.e. $g(z) = z$. We will first deal with the case $c = 0$. In this case

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

which means $z = \infty$ or $z = b/(d-a)$. Note that $b \neq 0$ since $c = 0$ and g is not the identity. Moreover, $ad = 1$ and so

$$\mathrm{tr}(g) = (a+d)^2 = (a+1/a)^2 = a^2 + 2 + 1/a^2 = 4 + (a^2 - 2 + 1/a^2) = 4 + (a - 1/a)^2 \geqslant 4$$

Moreover, $\operatorname{tr}(g) = 4$ if and only if $a = 1/a = d$ if and only if ∞ is the only fixed point.

We now turn to the $c \neq 0$ case. Here we have

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

which is equivalent to

$$z^2 + \left(\frac{d-a}{c}\right) z - \frac{b}{c} = 0.$$

This equation has two solutions in $\hat{\mathbb{C}}$ given by

$$\left(\frac{a-d}{2c}\right) \; \pm \; \frac{1}{c}\sqrt{\left(\frac{a-d}{2}\right)^2 + bc}.$$

Recall that c is real and non-zero. Writing

$$\Delta = \left(\frac{a-d}{2}\right)^2 + bc,$$

note that

$$\begin{aligned} \Delta \;=\; \frac{(a-d)^2}{4} + bc \;=\; \frac{a^2 - 2ad + d^2}{4} + bc \;&=\; \frac{a^2 + 2ad + d^2}{4} + bc - ad \\ &=\; \frac{(a+d)^2}{4} - 1 \\ &=\; \frac{\operatorname{tr}(g)^2}{4} - 1. \end{aligned}$$

We consider the three cases where Δ is positive, negative, and zero separately:

1. $\Delta > 0$. By the above, this is equivalent to $\operatorname{tr}(g)^2 > 4$ and so g is necessarily hyperbolic. It is also clear in this situation that g has precisely two fixed points in $\hat{\mathbb{C}}$, both of which lie in $\mathbb{R} \cup \{\infty\}$.

2. $\Delta = 0$. By the above, this is equivalent to $\operatorname{tr}(g)^2 = 4$ and so g is necessarily parabolic. It is also clear in this situation that g has precisely one fixed point in $\hat{\mathbb{C}}$, which lies in $\mathbb{R} \cup \{\infty\}$.

3. $\Delta < 0$. By the above, this is equivalent to $\operatorname{tr}(g)^2 < 4$ and so g is necessarily elliptic. It is also clear in this situation that g has precisely two fixed points which are complex conjugates and therefore one of which lies in $\mathbb{H}^2$ and the other in $\overline{\mathbb{H}^2}$.

This completes the proof. □

Fortunately, this classification also holds for $\operatorname{con}^+(1)$.

Lemma 4.2. *Let $g \in \operatorname{PSL}(2, \mathbb{R})$ and let $h \in \operatorname{con}^+(1)$ such that $g = \phi^{-1} \circ h \circ \phi$. Then*

$$\operatorname{tr}(g) = \operatorname{tr}(h).$$

Proof. This is a question on the tutorial sheets. □

4.1 Hyperbolic elements

The simplest example of a hyperbolic element is the dilation $z \mapsto \alpha z$ for a positive real number α not equal to 0 or 1. First note that this is indeed a hyperbolic element of $\mathrm{PSL}(2, \mathbb{R})$ since

$$\alpha z = \frac{\sqrt{\alpha} z + 0}{0 \times z + 1/\sqrt{\alpha}}$$

and there are two distinct fixed points in the boundary of $\mathbb{H}^2$ given by 0 and ∞. Moreover, if the fixed points of a hyperbolic element of $\mathrm{PSL}(2, \mathbb{R})$ are 0 and ∞, then it is necessarily of this form. If this is not clear, examine the proof of Theorem 4.1. Conversely, *any* hyperbolic element of $\mathrm{PSL}(2, \mathbb{R})$ is conjugate to a hyperbolic element of this simple form. In particular, suppose $g \in \mathrm{PSL}(2, \mathbb{R})$ is hyperbolic with distinct fixed points $v, w \in \mathbb{R} \cup \{\infty\}$. Then there exists an element $h_{v,w} \in \mathrm{PSL}(2, \mathbb{R})$ which sends v to 0 and w to ∞, i.e.

$$h_{v,w}(v) = 0 \qquad \text{and} \qquad h_{v,w}(w) = \infty.$$

First note that we can swap 0 and ∞ via the map $z \mapsto -1/z$ (which is in $\mathrm{PSL}(2, \mathbb{R})$). Therefore we can assume without loss of generality that $v < w$. First we can send v to 0 via the map $z \mapsto z - v$ (check that this is in $\mathrm{PSL}(2, \mathbb{R})$). If $w = \infty$, then we are done since this map fixes ∞. Otherwise, w has been sent to $w - v > 0$ and we can send this point to ∞ (whilst keeping 0 fixed) via the map

$$z \mapsto \frac{-z}{z - (w - v)}$$

(again, check that this is in $\mathrm{PSL}(2, \mathbb{R})$). We can conjugate g by $h_{v,w}$ to obtain a hyperbolic element $h_{v,w} \circ g \circ h_{v,w}^{-1} \in \mathrm{PSL}(2, \mathbb{R})$ which fixes 0 and ∞. This conjugate is the standard form of g. Since the fixed points of a hyperbolic element lie on the boundary of hyperbolic space, it necessarily moves every point of $\mathbb{H}^2$ (or $\mathbb{D}^2$). However, the geodesic ray joining the fixed points on the boundary is preserved.

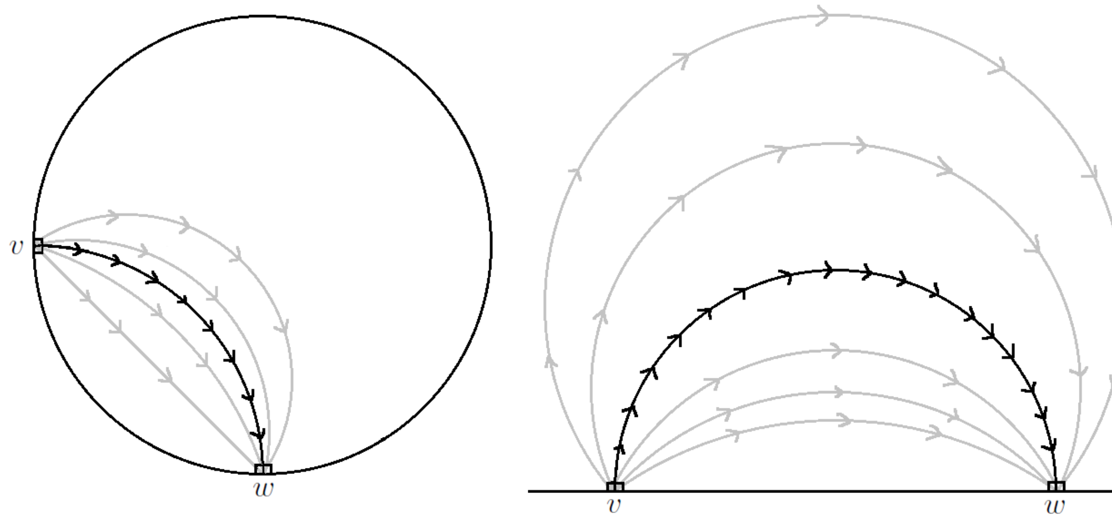
Lemma 4.3. *Let $g \in \mathrm{PSL}(2, \mathbb{R})$ be a hyperbolic element with fixed points $v, w \in \mathbb{R} \cup \{\infty\}$ and let C be the doubly infinite geodesic ray joining v and w. Then $g(C) = C$. Moreover, if $g \in \mathrm{PSL}(2, \mathbb{R})$ is not the identity and fixes a doubly infinite geodesic ray joining $v, w \in \mathbb{R} \cup \{\infty\}$, then g is a hyperbolic element with fixed points v, w.*

Proof. Let $h_{v,w} \circ g \circ h_{v,w}^{-1} \in \mathrm{PSL}(2, \mathbb{R})$. This is necessarily a hyperbolic map of the form $z \mapsto \alpha z$ for some $\alpha > 0$ not equal to 1. Moreover, $h_{v,w}$ maps C to the geodesic ray joining 0 and ∞, which is the straight vertical line above 0, and this line is clearly preserved by $z \mapsto \alpha z$. Therefore

$$C = h_{v,w}^{-1}(h_{v,w}(C)) = h_{v,w}^{-1}(h_{v,w}(g(h_{v,w}^{-1}(h_{v,w}(C))))) = g(C)$$

as required. The converse result is immediate since a map which fixes a doubly infinite geodesic ray necessarily fixes its endpoints (since Möbius maps are continuous) and if an element of $\mathrm{PSL}(2, \mathbb{R})$ has two fixed points, then it is hyperbolic. $\square$

Again by considering the standard form of a hyperbolic element, we see that one of the fixed points is attracting and one is repelling. Let $g \in \mathrm{PSL}(2, \mathbb{R})$ be a hyperbolic element in standard form, i.e. given by $z \mapsto \alpha z$ for some $\alpha > 0$ not equal to 1. If $\alpha \in (0, 1)$, then all points in $\mathbb{H}^2$ are attracted to 0 (and repelled from ∞) and if $\alpha > 0$, then all points in $\mathbb{H}^2$ are repelled from 0 (and attracted to ∞).



4.2 Parabolic elements

The simplest example of a parabolic element is the translation $z \mapsto z + \beta$ for a non-zero real number β. Again, this is indeed a parabolic element of $\mathrm{PSL}(2, \mathbb{R})$ and the only fixed point is ∞. Moreover, if a parabolic element of $\mathrm{PSL}(2, \mathbb{R})$ fixes ∞, then it is necessarily of this form. Conversely, *any* parabolic element of $\mathrm{PSL}(2, \mathbb{R})$ is conjugate to a parabolic element of this simple form. In particular, suppose $g \in \mathrm{PSL}(2, \mathbb{R})$ is parabolic with fixed point $v \in \mathbb{R}$. Then there exists an element $h_v \in \mathrm{PSL}(2, \mathbb{R})$ which sends v to ∞. Indeed, the map

$$z \mapsto \frac{-1}{z - v}$$

does the job. We can conjugate g by h_v to obtain a parabolic element $h_v \circ g \circ h_v^{-1} \in \mathrm{PSL}(2, \mathbb{R})$ which fixes ∞. This conjugate is the standard form of g. Since the fixed point of a parabolic element lies on the boundary of hyperbolic space, it necessarily moves every point of $\mathbb{H}^2$ (or $\mathbb{D}^2$). However, it does preserve any *horocycle* which passes through the fixed point. A horocycle is a circle which is tangent to the boundary of $\mathbb{H}^2$. This either takes the form of a doubly infinite horizontal Euclidean straight line or a circle which is touches $\mathbb{R}$ at one point.

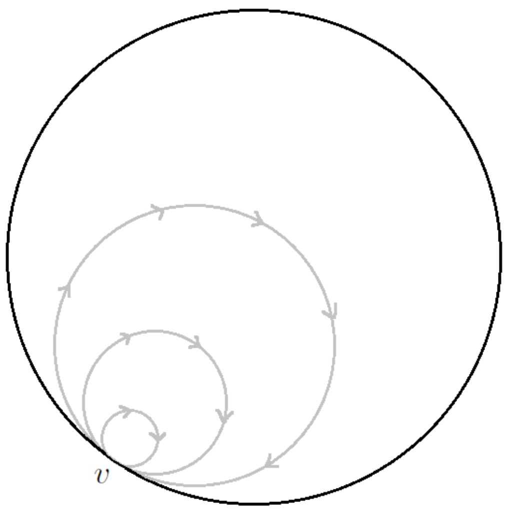
Lemma 4.4. *Let $g \in \mathrm{PSL}(2, \mathbb{R})$ be a parabolic element with fixed point $v \in \mathbb{R} \cup \{\infty\}$ and let H be any horocycle passing through v. Then $g(H) = H$. Moreover, if $g \in \mathrm{PSL}(2, \mathbb{R})$ is not the identity and fixes a horocycle, then g is a parabolic element with fixed point equal to the base point of the horocycle.*

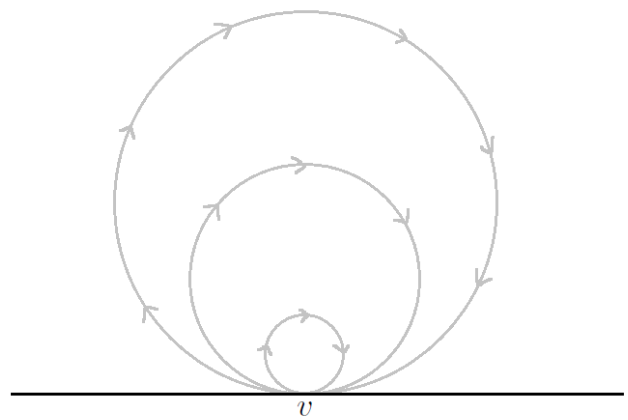
Proof. Let $h_v \circ g \circ h_v^{-1} \in \mathrm{PSL}(2, \mathbb{R})$. This is necessarily a parabolic map of the form $z \mapsto z + \beta$ for some $\beta \in \mathbb{R}$. Moreover, h_v maps H to a doubly infinite horizontal line, and this line is clearly preserved by $z \mapsto z + \beta$. Therefore

$$H = h_v^{-1}(h_v(H)) = h_v^{-1}(h_v(g(h_v^{-1}(h_v(H))))) = g(H)$$

as required. The converse result follows since any non-identity map which fixes a horocycle also fixes its base point on the boundary and is therefore hyperbolic or parabolic. However, if it was hyperbolic it would also fix a geodesic ray between the base point of the horocycle and the other fixed point on the boundary. This geodesic ray meets the horocycle at one point in the $\mathbb{H}^2$ which is therefore fixed by the map. This is a contradiction. Put differently, a hyperbolic element cannot fix a horocycle. $\square$

Again by considering the standard form of a parabolic element, we see that its fixed point is both attracting and repelling and the orbit of a given point lies on a horocycle.





4.3 Elliptic elements

The simplest example of an elliptic element is the rotation

$$z \mapsto \frac{\cos\theta z + \sin\theta}{-\sin\theta z + \cos\theta}$$

for an angle $\theta \in (0, \pi)$. This is indeed an elliptic element since $\mathrm{tr}(g)^2 = (2\cos\theta)^2 < 4$. The (conjugate pair) of fixed points are $\pm i$ and the action of this map on $\mathbb{H}^2$ should be thought of as 'hyperbolic rotation' around i. Indeed the conjugation of an elliptic element in standard form by ϕ is Euclidean rotation around 0 in $\mathbb{D}^2$. It is a little less obvious this time that if an elliptic element of $\mathrm{PSL}(2, \mathbb{R})$ fixes i, then it is necessarily of this form.

Lemma 4.5. *Let $g \in \mathrm{PSL}(2, \mathbb{R})$ be an elliptic element which fixes i. Then g has the standard form stated above.*

Proof. Suppose that

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

for $a, b, c, d \in \mathbb{R}$ with $ad - bc = 1$ and

$$i = g(i) = \frac{ai + b}{ci + d}$$

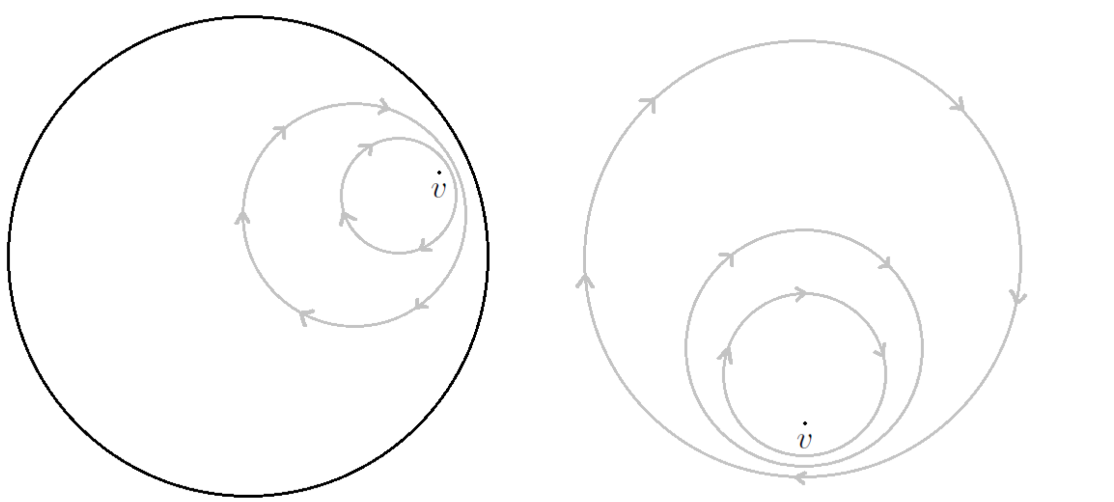
which means $b + ai = -c + di$ and therefore $a = d$ and $b = -c$. Combining this with $ad - bc = 1$, we obtain $a^2 + b^2 = 1$ and $a^2 + c^2 = 1$, and hence $d = a, b = \pm\sqrt{1 - a^2}, c = \mp\sqrt{1 - a^2}$. Using

$ad - bc = 1$ again, we see that b and c must have opposite signs. In the case of b positive and c negative we may set $a = \cos\theta$ to obtain the desired conclusion. On the other hand, if we take b negative and c positive we may set $a = \cos(-\theta)$ and use the fact that cos is an even function and sin is odd. □

Similar to the hyperbolic and parabolic case, *any* elliptic element of $\mathrm{PSL}(2,\mathbb{R})$ is a conjugate of an elliptic element of this simple form. This time we conjugate by the map which sends the fixed point (in $\mathbb{H}^2$) of a given elliptic element to i and then apply the above lemma. We have already seen that we can send any point in the interior of $\mathbb{H}^2$ to any other point via an element of $\mathrm{PSL}(2,\mathbb{R})$. Contrary to the other cases, elliptic elements fix (precisely one) point in $\mathbb{H}^2$. They also leave hyperbolic circles centred at the fixed point invariant.

Lemma 4.6. *Let $g \in \mathrm{PSL}(2,\mathbb{R})$ be an elliptic element with fixed point $v \in \mathbb{H}^2$ and let C be any hyperbolic circle centred at v. Then $g(C) = C$. Moreover, if $g \in \mathrm{PSL}(2,\mathbb{R})$ is not the identity and fixes a hyperbolic circle, then g is an elliptic element with fixed point equal to the centre of the circle.*

Proof. Since g is a Möbius map, we know that $g(C)$ is a hyperbolic circle. Moreover, since the centre is fixed by g, we can guarantee that $g(C) = C$. In the converse direction, if g fixes a hyperbolic circle, it necessarily fixes its centre (since g is an isometry) and if a non-identity element fixes a point in the interior of $\mathbb{H}^2$, then it must be elliptic. □



5 Fuchsian groups

Fuchsian groups are an important class of groups of isometries of hyperbolic space, i.e. subgroups of the group of isometries. Before we define them, we need to recall the notion of *discreteness.*

Definition 5.1. *Let (X, d) be a metric space and $E \subseteq X$ a subset. We say E is* discrete *if for all $x \in E$, there exists $r > 0$ such that*

$$B(x, r) \cap E = \{x\}.$$

The intuitive way to think of a discrete set is that all points are isolated or, equivalently, there are no accumulation points. It does not alter the above definition if we take the ball $B(x, r)$ to be open or closed, but it is more traditional to think of it as being open.

We can turn $\mathrm{PSL}(2, \mathbb{R})$ into a metric space simply by viewing it as a (3 dimensional) submanifold of $\mathbb{R}^4$. In particular, for $g \in \mathrm{PSL}(2, \mathbb{R})$ given by

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

for $a, b, c, d \in \mathbb{R}$ with $ad - bc = 1$, we define the norm of g to be the Euclidean norm of the vector (a, b, c, d), i.e.

$$\|g\| = |(a, b, c, d)| = \sqrt{a^2 + b^2 + c^2 + d^2}.$$

Recall that normed spaces are metric spaces. Fuchsian groups are precisely the discrete subgroups of $\mathrm{PSL}(2, \mathbb{R})$.

Definition 5.2. *A group $\Gamma \leqslant \mathrm{PSL}(2, \mathbb{R})$ is a* Fuchsian group *if it is a discrete subset of $\mathrm{PSL}(2, \mathbb{R})$ equipped with the Euclidean norm.*

We can also define Fuchsian subgroups of $\mathrm{con}^+(1)$ in a similar way, this time identifying $\mathrm{con}^+(1)$ with a 3 dimensional submanifold of $\mathbb{C}^2$. It is straightforward to show that $\Gamma \leqslant \mathrm{con}^+(1)$ is Fuchsian if and only if $\phi^{-1}\Gamma\phi \leqslant \mathrm{PSL}(2, \mathbb{R})$ is Fuchsian.

The archetypal example of a Fuchsian group is the *modular group* $\mathrm{PSL}(2, \mathbb{Z})$ which is the subgroup of $\mathrm{PSL}(2, \mathbb{R})$ where $a, b, c, d \in \mathbb{Z}$. It turns out that Fuchsian groups act on $\mathbb{H}^2$ in a particularly elegant way, known as *properly discontinuously.* This type of action leads to natural 'tilings' of hyperbolic space, and limit sets with interesting geometrical properties.

Definition 5.3. *A group G acting on a metric space X is said to act* properly discontinuously *if the* orbit

$$G(x) = \{g(x) : g \in G\}$$

is locally finite *for all $x \in X$. (A set $E \subseteq X$ is called* locally finite *if for all compact sets $K \subseteq X$ the set $K \cap E$ is finite.)*

The following important theorem gives another (more geometric/dynamic) way to classify Fuchsian groups.

Theorem 5.4. *A group $\Gamma \leqslant \mathrm{PSL}(2, \mathbb{R})$ is Fuchsian if and only if it acts properly discontinuously on $\mathbb{H}^2$.*

Proof. ($\Rightarrow$). Let Γ be a Fuchsian group, and therefore discrete. Let $z \in \mathbb{H}^2$ and $K \subseteq \mathbb{H}^2$ be compact, and therefore closed and bounded. To show Γ acts properly discontinuously, it suffices to show that

$$|\Gamma(z) \cap K| \leqslant |\Gamma \cap \{g \in \mathrm{PSL}(2,\mathbb{R}) : g(z) \in K\}| < \infty.$$

The first inequality is clear and, since Γ is discrete and closed, it suffices to show that $\{g \in \mathrm{PSL}(2,\mathbb{R}) : g(z) \in K\}$ is compact since the intersection of a closed discrete set with a compact set is finite. (It is an exercise in the tutorial sheet to prove that Fuchsian groups are closed and that the intersection of a closed discrete set with a compact set is at most finite).

To see that $\{g \in \mathrm{PSL}(2,\mathbb{R}) : g(z) \in K\}$ is closed, let $g_n \in \mathrm{PSL}(2,\mathbb{R})$ be such that $g_n(z) \in K$ for all n and $g_n \to g \in \mathrm{PSL}(2,\mathbb{R})$. Then $g(z) = \lim_{n\to\infty} g_n(z) \in K$ since K is closed and each of $g_n(z) \in K$.

We will now show that $\{g \in \mathrm{PSL}(2,\mathbb{R}) : g(z) \in K\}$ is bounded, and thus compact. It suffices to prove that there is a uniform bound on a, b, c, d satisfying

$$\frac{az+b}{cz+d} \in K.$$

Since K is bounded in $\mathbb{H}^2$, this means that there exists a constant $C > 1$ such that for all $w \in K$ we have $|w| \leqslant C$ and $\mathrm{Im}(w) > 1/C$. Therefore

$$|az+b| \leqslant C|cz+d|$$

and

$$1/C < \mathrm{Im}\left(\frac{az+b}{cz+d}\right) = \mathrm{Im}\left(\frac{(az+b)(c\overline{z}+d)}{|cz+d|^2}\right) = \frac{\mathrm{Im}(adz+bc\overline{z})}{|cz+d|^2} = \frac{\mathrm{Im}(z)(ad-bc)}{|cz+d|^2} = \frac{\mathrm{Im}(z)}{|cz+d|^2}$$

and therefore

$$|cz+d| \leqslant \sqrt{C\mathrm{Im}(z)}$$

which means both $|az+b|$ and $|cz+d|$ are bounded. We conclude that a, b, c, d are all individually bounded, as required. For example $|az+b| \geqslant |a|\mathrm{Im}(z)$ and so a is bounded and also $|az+b| \geqslant |a\mathrm{Re}(z)+b|$ and so b is bounded.

($\Leftarrow$). Suppose $\Gamma \leqslant \mathrm{PSL}(2,\mathbb{R})$ acts properly discontinuously on $\mathbb{H}^2$. Assume that Γ is not discrete and therefore we can find an element $g \in \Gamma$ and a sequence $g_n \in \Gamma \setminus \{g\}$ such that $g_n \to g$. Let $h_n = g^{-1} \circ g_n \in \Gamma$ and note that $h_n \neq \mathrm{Id}$ for all n, but $h_n \to \mathrm{Id}$. We now have two cases, at least one of which must hold:

1. Infinitely many of the h_n fix i. In this case we can find an infinite sequence of (non-identity) elliptic elements which all fix i and converge to the identity. Consider the orbit $\Gamma(2i)$ and in particular the image of $2i$ under this sequence of elliptic elements. This cannot be locally finite because $2i$ is an accumulation point.

2. Infinitely many of the h_n do not fix i. In this case i is an accumulation point of the orbit $\Gamma(i)$, again contradicting the fact that Γ acts properly discontinuously.

We deduce that Γ must be discrete, and thus Fuchsian, completing the proof. $\square$

5.1 Fundamental domains

We mentioned earlier that Fuchsian groups lead to tilings of hyperbolic space. One should think of a fundamental domain as a tile for a given Fuchsian group, i.e. a set whose orbit is a tiling.

Definition 5.5. *Let $\Gamma \leqslant \mathrm{PSL}(2,\mathbb{R})$ be a Fuchsian group. A* fundamental domain *for Γ is an open set $F \subseteq \mathbb{H}^2$ such that*

1. *the whole space is 'tiled', i.e. $\mathbb{H}^2 = \bigcup_{g\in\Gamma} g\left(\overline{F}\right)$*
2. *the 'tiles' don't overlap, i.e. for all $g, h \in \Gamma$ with $g \neq h$, we have $g(F) \cap h(F) = \emptyset$.*

Note that $\overline{F}$ denotes the closure of F in the hyperbolic metric, although this is the same as the Euclidean closure, except at the boundary! Our first task is to show that fundamental domains always exist and to give a canonical way to build one.

Definition 5.6. *Let $\Gamma \leqslant \mathrm{PSL}(2,\mathbb{R})$ be a Fuchsian group and $z \in \mathbb{H}^2$ be a point not fixed by any element of Γ. Then the* Dirichlet region *of Γ at z is given by*

$$D_z(\Gamma) = \bigcap_{g\in\Gamma\setminus\{\mathrm{Id}\}} \left\{ w \in \mathbb{H}^2 : d_{\mathbb{H}^2}(w,z) < d_{\mathbb{H}^2}(w,g(z)) \right\}.$$

Theorem 5.7. *Let $\Gamma \leqslant \mathrm{PSL}(2,\mathbb{R})$ be a Fuchsian group. Any Dirichlet region for Γ is a connected convex fundamental domain.*

Proof. First note that $D_z(\Gamma)$ is the intersection of convex sets (half-spaces) and is therefore convex and connected. $D_z(\Gamma)$ is also a (usually infinite) intersection of open sets, which in general need not be open. However, openness follows the fact that Γ acts properly discontinuously. Let $u \in D_z(\Gamma)$ and $B(u,r) \subset \mathbb{H}^2$ be the (closed) hyperbolic ball centred at u with radius r. Since Γ acts properly discontinuously, $|\Gamma(z) \cap B(u, 5d_{\mathbb{H}^2}(u,z))| < \infty$ and so for all but finitely many $g \in \Gamma$ we have

$$B(u, d_{\mathbb{H}^2}(u,z)) \cap \left\{ w \in \mathbb{H}^2 : d_{\mathbb{H}^2}(w,z) < d_{\mathbb{H}^2}(w,g(z)) \right\} = B(u, d_{\mathbb{H}^2}(u,z)).$$

Therefore we can find a neighbourhood of u inside $D_z(\Gamma)$ by taking the *finite* intersection of $B(u, d_{\mathbb{H}^2}(u,z))$ with the half-spaces corresponding to g which fail the above.

To show that $D_z(\Gamma)$ is a fundamental domain we must prove the two conditions from the definition.

1. Let $w \in \mathbb{H}^2$ be given. Since $\Gamma(w)$ is a closed discrete set (see tutorial questions), there exists $h \in \Gamma$ which minimises the distance of $\Gamma(w)$ to z. That is $d_{\mathbb{H}^2}(h(w),z) \leqslant d_{\mathbb{H}^2}(g(w),z)$ for all $g \in \Gamma$. Therefore for all $g \in \Gamma$ we have

$$d_{\mathbb{H}^2}(h(w),z) \leqslant d_{\mathbb{H}^2}(w, g^{-1}(z)) = d_{\mathbb{H}^2}(h(w), h(g^{-1}(z))).$$

Since arbitrary elements of Γ can be written in the form $h \circ g^{-1}$ this proved that

$$h(w) \in \overline{D_z(\Gamma)}$$

and so

$$w \in h^{-1}(\overline{D_z(\Gamma)}) \subseteq \bigcup_{g\in\Gamma} g\left(\overline{D_z(\Gamma)}\right)$$

as required.

2. Let $w_1, w_2 \in \Gamma(v)$ for some $v \in \mathbb{H}^2$ be such that $w_1 \neq w_2$ and assume that $w_1 \in D_z(\Gamma)$. We will first prove that $w_2 \notin D_z(\Gamma)$. Let $g_1, g_2 \in \Gamma$ be such that $w_1 = g_1(v)$ and $w_2 = g_2(v)$. Since $w_1 \in D_z(\Gamma)$ we have

$$d_{\mathbb{H}^2}(w_1, z) < d_{\mathbb{H}^2}(w_1, g(z))$$

for all $g \in \Gamma \setminus \{\mathrm{Id}\}$. By choosing $g = g_1 g_2^{-1}$, we obtain

$$d_{\mathbb{H}^2}(w_1, z) < d_{\mathbb{H}^2}(w_1, g_1(g_2^{-1}(z))) = d_{\mathbb{H}^2}(g_2(g_1^{-1}(w_1)), z) = d_{\mathbb{H}^2}(w_2, z).$$

Then note that

$$d_{\mathbb{H}^2}(w_1, z) = d_{\mathbb{H}^2}(g_1(v), z) = d_{\mathbb{H}^2}(v, g_1^{-1}(z)) = d_{\mathbb{H}^2}(g_2(v), g_2(g_1^{-1}(z))) = d_{\mathbb{H}^2}(w_2, g_2(g_1^{-1}(z))).$$

Therefore

$$d_{\mathbb{H}^2}(w_2, z) > d_{\mathbb{H}^2}(w_2, g_2(g_1^{-1}(z))),$$

which implies $w_2 \notin D_z(\Gamma)$.

To complete the proof, suppose $v \in g_1(D_z(\Gamma)) \cap g_2(D_z(\Gamma))$ for some $g_1, g_2 \in \Gamma$ with $g_1 \neq g_2$. Since $g_1(D_z(\Gamma)) \cap g_2(D_z(\Gamma))$ is open, we can choose $r > 0$ sufficiently small to guarantee that $B_E(v, r) \subseteq g_1(D_z(\Gamma)) \cap g_2(D_z(\Gamma))$. Then for any point $u \in B_E(v, r)$ we have $g_1^{-1}(u), g_2^{-1}(u) \in D_z(\Gamma)$, which by the above argument implies $g_1^{-1}(u) = g_2^{-1}(u)$. Thus we have proved that $g_1 = g_2$ on an open set $g_1^{-1}(B_E(v, r))$, which proves that $g_1 = g_2$.

The proof is complete. □

From now on we will tend to assume that Dirichlet fundamental domains are polygonal, by which we mean that they are convex hyperbolic polygons with a finite number of vertices in $\overline{\mathbb{H}^2}$ and a finite number of edges, each of which is a geodesic segment or a geodesic ray in $\mathbb{H}^2$.

Lemma 5.8. *Let $D_z(\Gamma)$ be a polygonal Dirichlet fundamental domain for a Fuchsian group Γ. Label the edges which bound $D_z(\Gamma)$ by e_g where $g \in \Gamma$ is such that e_g is part of the perpendicular bisector of the geodesic between z and $g(z)$. Then*

$$g^{-1}(e_g) = e_{g^{-1}} \ \text{ and } \ g(e_{g^{-1}}) = e_g.$$

Proof. Let

$$H_g = \left\{ w \in \mathbb{H}^2 : d_{\mathbb{H}^2}(w, z) = d_{\mathbb{H}^2}(w, g(z)) \right\}$$

be the geodesic ray containing e_g. Observe that

$$\begin{aligned} w \in H_g \quad &\Longleftrightarrow \quad d_{\mathbb{H}^2}(w, z) = d_{\mathbb{H}^2}(w, g(z)) \\ &\Longleftrightarrow \quad d_{\mathbb{H}^2}(g^{-1}(w), g^{-1}(z)) = d_{\mathbb{H}^2}(g^{-1}(w), z) \\ &\Longleftrightarrow \quad g^{-1}(w) \in H_{g^{-1}} \end{aligned}$$

which shows that $g^{-1}(H_g) = H_{g^{-1}}$. A little more work allows us to conclude that $g^{-1}(e_g) = e_{g^{-1}}$, but we omit the details. One way to prove it is to note that the points $u \in H_g$ are in one to one correspondence with the (clockwise) angles $\theta \in (0, 2\pi)$ formed between the geodesic ray H_g and the geodesic from z to u and that g is conformal and so must preserve these angles. Draw a picture! □

Elements of a Fuchsian group which take an edge of a fundamental domain to another edge are called called *side-pairing transformations*. The above lemma implies that the sides of a polygonal Dirichlet fundamental domain come in pairs, which are mapped to each other by (unique) members of the Fuchsian group. There is one exception to this, which occurs if there is an elliptic fixed point contained in one of the sides. In this case the elliptic transformation maps this side to itself: fixing the fixed point and interchanging the segments either side of it. Therefore, if we have such an elliptic fixed point, then we view it as a vertex of the fundamental domain and the elliptic transformation pairs the sides meeting at this vertex. With this convention, if $D_z(\Gamma)$ is a polygonal Dirichlet fundamental domain then the sides come in pairs which are mapped onto each other by the side-pairing transformations. This means that the number of sides is always even!

Theorem 5.9. *The collection of all the side-pairing transformations of a Dirichlet fundamental domain for a Fuchsian group* Γ *are a generating set for* Γ*.*

Proof. Let D be a Dirichlet fundamental domain for Γ, and let H be the group generated by the side-pairing transformations. Clearly $H \leqslant \Gamma$ and so the aim is to show that $\Gamma \subseteq H$. First note that if $h \in H$ and $g \in \Gamma$ are such that $h(D)$ and $g(D)$ are adjacent (i.e. have one edge in common), then necessarily $g \in H$. This follows since $h^{-1}h(D) = D$ and $h^{-1}g(D)$ are adjacent, which implies that $h^{-1}g$ is a side-pairing transformation, and hence $h^{-1}g \in H$. Then using the fact that H is a group we know $g = hh^{-1}g \in H$ as required.

Now, let $g \in \Gamma$ be given with the aim to show that $g \in H$. In fact the above observation already does the job via an inductive argument. Since $\Gamma(\overline{D})$ is a 'tiling' of the hyperbolic plane, we can find a 'chain' of images $g_1(D), g_2(D), \ldots, g_n(D)$ where $g_1 = \mathrm{Id}$, $g_n = g$, and $g_k(D)$ is adjacent to $g_{k+1}(D)$ for $k \in \{1, \ldots, n-1\}$. Therefore for each such k, we have $g_k \in H \Rightarrow g_{k+1} \in H$, and the desired result follows by induction. □

The above theorem shows that the sides of a polygonal Dirichlet fundamental domain are in one-to-one correspondence with a set of *generators* for the Fuchsian group. It turns out that the vertices of the fundamental domain are in one-to-one correspondence with the *relations* in a given presentation for the group.

We briefly recall the notion of a presentation. Given a finite set A and a finite set $B \subset F_A$ where F_A is the free group over A, then we write

$$H \cong \langle A : B \rangle$$

to mean the group generated by the elements of A, subject to the constraints that $b = \mathrm{Id}$ for all $b \in B$. This way of expressing a group is known as a *presentation* and since A, B are finite we say H is *finitely presented*. Elements of A are called *generators* and elements of B are called *relations*. We will not worry about the details of presentations here but, more formally,

$$\langle A : B \rangle \cong F_A / \left\langle B^{F_A} \right\rangle$$

where B^{F_A} is the conjugation of B by F_A, i.e.

$$B^{F_A} = \{g^{-1}hg : g \in F_A, h \in B\}.$$

This generates a normal subgroup $\left\langle B^{F_A} \right\rangle \trianglelefteq F_A$ known as the *normal closure* of B in F_A. Thus the presentation $\langle A : B \rangle$ is formally the quotient of the free group over A by the normal closure of B. Examples include, free groups

$$F_A \cong \langle A : \emptyset \rangle ,$$

cyclic groups of order n

$$C_n \cong \langle a : a^n \rangle ,$$

and the integer lattice (or free abelian group of order 2)

$$\mathbb{Z} \times \mathbb{Z} \cong \left\langle a, b : a^{-1}b^{-1}ab \right\rangle .$$

Let E denote the finite set of edges and V the finite set of vertices of a polygonal Dirichlet fundamental domain. For each $v \in V \cap \mathbb{H}^2$ we associate a relation as follows:

Step 1. Let $v_1 = v \in V$, $e_{g_1} \in E$ be one of the edges adjacent to v_1, and $\alpha_1(v)$ be the angle at v_1. Therefore $g_1^{-1}e_{g_1} = e_{g_1^{-1}} \in E$ and $g_1^{-1}(v_1) \in V$ is a vertex adjacent to $e_{g_1^{-1}}$.

Step 2. Let $v_2 = g_1^{-1}(v_1)$, and $\alpha_2(v)$ be the angle at v_2. Since the domain is polygonal, there is a unique edge $e_{g_2} \in E$ different from $e_{g_1^{-1}}$ adjacent to v_2. Therefore $g_2^{-1}e_{g_2} = e_{g_2^{-1}} \in E$ and $g_2^{-1}(v_2) \in V$ is a vertex adjacent to $e_{g_2^{-1}}$.

Step 3. Let $v_3 = g_2^{-1}(v_2) = g_2^{-1}(g_1^{-1}(v_1))$, and $\alpha_3(v)$ be the angle at v_3. Since the domain is polygonal, there is a unique edge $e_{g_3} \in E$ different from $e_{g_2^{-1}}$ adjacent to v_3. Therefore $g_3^{-1}e_{g_3} = e_{g_3^{-1}} \in E$ and $g_3^{-1}(v_3) \in V$ is a vertex adjacent to $e_{g_3^{-1}}$.

$\vdots$

Step $k(v)$. Since V is finite, this process must return to $v = v_1$ in a finite number $k(v)$ of steps. This does not quite follow directly from finiteness (this just gives that we eventually must reach a vertex we have seen before), but the only vertex one can possibly visit twice is v_1. This is because every vertex is adjacent to two sides and therefore has two paths to it via this process. Once a vertex has been visited, both of these paths have been used (one 'in' and one 'out') and since we only went 'out' of v_1, we must eventually go back 'in' to it.

Once we have returned to v_1 we terminate the algorithm and conclude that

$$v = g_{k(v)}^{-1}(g_{k(v)-1}^{-1}(\cdots g_2^{-1}(g_1^{-1}(v))\cdots)).$$

This means that the *cycle transformation at* v given by $g_1 \cdots g_{k(v)}$ fixes $v \in \mathbb{H}^2$ and so is either equal to the identity, or is an elliptic element with finite order $n(v)$, i.e. for some (minimal) integer $n(v) \geqslant 1$ we have

$$(g_1 \cdots g_{k(v)})^{n(v)} = \mathrm{Id}.$$

Note that an elliptic element of a Fuchsian group must have finite order, or it would violate the fact that the group acts properly discontinuously.

If the cycle transformation is elliptic, then the orbit of the Dirichlet fundamental domain under the the action of this map tiles a neighbourhood of v and this implies that v satisfies the *angle sum condition*:

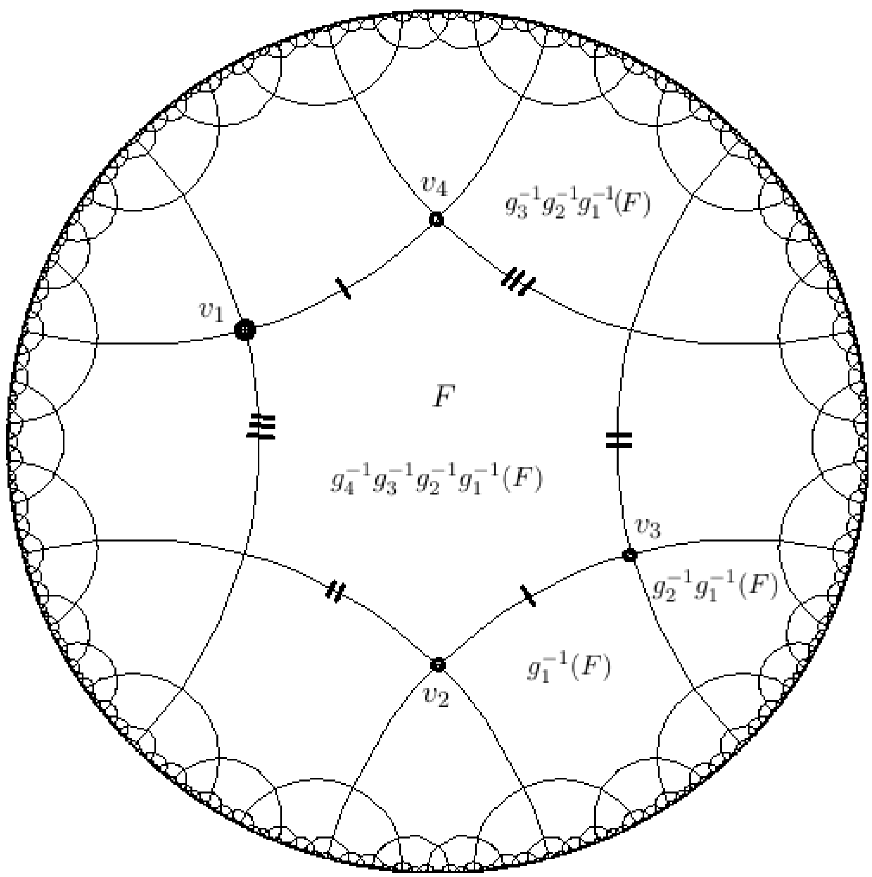
$$n(v) \sum_{i=1}^{k(v)} \alpha_i(v) = 2\pi.$$

We have almost proved the following theorem.

Theorem 5.10. *Suppose Γ is a Fuchsian group and $D_z(\Gamma)$ is a polygonal Dirichlet fundamental domain with finite vertex set V. Let $\mathcal{E}$ denote the complete set of side-pairing transformations for $D_z(\Gamma)$ and for $v \in V$ let g_v be the associated cycle transformation. Then*

$$\Gamma \;\cong\; \left\langle \mathcal{E} \;:\; g_v^{n(v)}, v \in V \right\rangle .$$

Proof. What we have actually proved is that Γ is isomorphic to a quotient of $\left\langle \mathcal{E} \;:\; g_v^{n(v)}, v \in V \right\rangle$ by the kernal of an appropriate homomorphism. To complete the proof one can show that the kernal is trivial, but we omit the details. A full proof can be found in Katok, for example. □



The above figure shows an example of the side-pairing algorithm for detecting relations. The paired sides of the fundamental domain F are indicated with dashes and, beginning with the point marked v_1, we find the algorithm terminates after 4 steps and $g_4^{-1}g_3^{-1}g_2^{-1}g_1^{-1}(F) = F$. Therefore, $g_1g_2g_3g_4$ is a relation in the induced presentation of the underlying Fuchsian group given by Theorem 5.10.

5.2 Poincaré's Theorem

The last result in the previous section allows us to take a group, move to a fundamental domain, and then recover the group from the geometry of the fundamental domain via a presentation. In this mini-section, we state a remarkable converse. It says that, given a sensible polygon (satisfying the angle sum condition at each vertex for a given set of side-pairing transformations), then one can obtain a Fuchsian group with this polygon as a fundamental domain.

If there are multiple vertices on the boundary of hyperbolic space, then there is another technical condition to worry about. In particular, imagine that v a boundary vertex. We can associate a cycle transformation with v using the same algorithm as before. Side-pairing transformations map boundary vertices to other boundary vertices and so if there is only one boundary vertex then the associated cycle transformation is either one of the side-pairing transformations associated to sides adjacent to v (these are mutually inverse). If there are multiple boundary vertices, then the situation is more complicated.

Theorem 5.11 (Poincaré's Theorem)**.** *Let P be a hyperbolic polygon and let $\mathcal{E}$ be a complete set of side-pairing transformations for P. Assume that every vertex satisfies the angle sum condition, i.e. for any vertex v of P, we have*

$$n(v)\sum_{i=1}^{k(v)} \alpha_i(v) = 2\pi$$

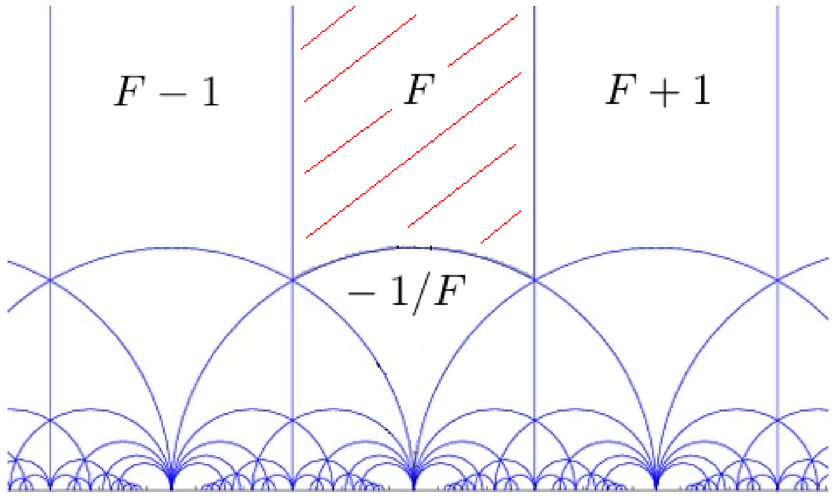
and that all cycle transformations associated with boundary vertices are either parabolic or the identity (i.e., not hyperbolic). Then the group generated by $\mathcal{E}$ is a Fuchsian group with P as a fundamental domain.

5.3 The modular group: fundamental domain, generators, and presentations

There does not seem to be a simple general method for computing the Dirichlet fundamental domain for a given Fuchsian group. A good strategy is to think of a generating set which corresponds to the side-pairing transformations of a convex polygon and then prove directly that this convex polygon is a Dirichlet region. As an example, we compute the Dirichlet fundamental domain for the modular group $\Gamma = \mathrm{PSL}(2,\mathbb{Z})$. Let

$$F = \left\{ w \in \mathbb{H}^2 \ : \ -1/2 < \mathrm{Re}(w) < 1/2 \text{ and } |z| > 1 \right\}.$$

We will prove that F is the Dirichlet region with base point $z = 2i$, i.e. $F = D_{2i}(\Gamma)$. First observe that $2i$ is not an elliptic fixed point of Γ. Indeed, recalling when we classified isometries by fixed points, the only way for $2i$ to be a fixed point of an element of $\mathrm{PSL}(2,\mathbb{R})$ was to choose $a = d$ and $b = -4c \neq 0$ which violates $ad - bc = a^2 + 4c^2 = 1$ since c is a non-zero integer. The parabolic map $z \mapsto z + 1$ is a side-pairing transformation for the vertical sides of F and the elliptic transformation $z \mapsto -1/z$ fixes i, splits the final side at i and pairs these sides. In particular, $F \supseteq D_{2i}(\Gamma)$.



We prove the reverse inclusion by contradiction. Suppose $D_{2i}(\Gamma)$ is a strict subset of F, which guarantees the existence of $w \in F$ and $g \in \Gamma$ such that $g(w) \in F$. This follows since $\Gamma(D_{2i}(\Gamma))$ tiles $\mathbb{H}^2$. Suppose g is defined by

$$g(w) = \frac{aw + b}{cw + d}$$

for $a, b, c, d \in \mathbb{Z}$ with $ad - bc = 1$. First note that $c \neq 0$, since if $c = 0$, then $g(w) = w + b$ and since b is an integer we cannot have both w and $g(w)$ inside F. It follows that

$$|cw + d|^2 = c^2|w|^2 + 2cd\,\mathrm{Re}(w) + d^2 > c^2 - |cd| + d^2 = (|c| - |d|)^2 + |cd| \geqslant 1$$

(in the final line we use that $c \neq 0$ and c and d are integers). It follows that $|cw + d|^2 > 1$ and hence

$$\mathrm{Im}(g(w)) = \mathrm{Im}\left(\frac{(aw + b)(c\overline{w} + d)}{|cw + d|^2}\right) = \frac{\mathrm{Im}(w)(ad - bc)}{|cw + d|^2} = \frac{\mathrm{Im}(w)}{|cw + d|^2} < \mathrm{Im}(w).$$

Arguing similarly with w and g replaced by $g(w)$ and g^{-1} we obtain

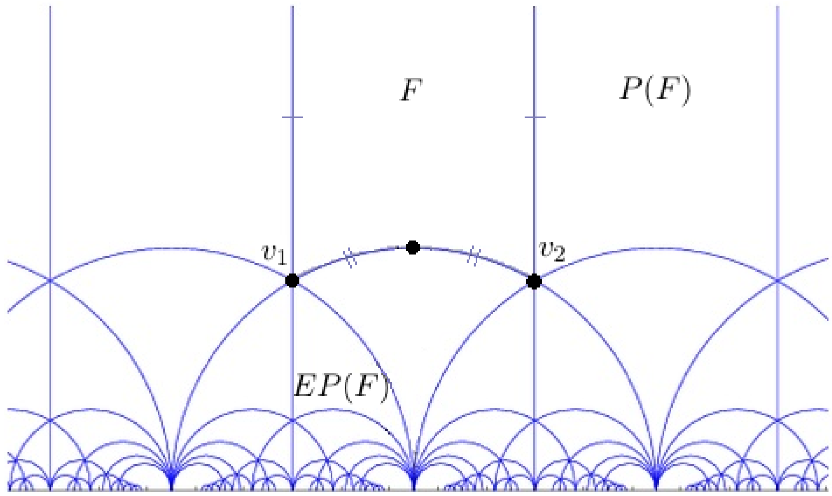
$$\mathrm{Im}(w) = \mathrm{Im}(g^{-1}(g(w))) < \mathrm{Im}(g(w)),$$

which is a contradiction as required. We obtain the following group theoretic corollary by applying Theorem 5.10.

Corollary 5.12. *The modular group* $\mathrm{PSL}(2, \mathbb{Z})$ *is generated by the parabolic map* $P : z \mapsto z+1$ *and the elliptic map* $E : z \mapsto -1/z$ *and, moreover, has the following presentation*

$$\mathrm{PSL}(2, \mathbb{Z}) \;\cong\; \left\langle P, E \;:\; E^2, (EP)^3 \right\rangle.$$

Proof. Since F is a Dirichlet fundamental domain it is generated by its side-pairing transformations $\{E, P, P^{-1}\}$ (and we may clearly drop P^{-1} from the generating set). Moreover, applying the side-pairing algorithm to the vertex i we find $E^2 = \mathrm{Id}$ and so E^2 is a relation and applying the algorithm to either of the other vertices in $\mathbb{H}^2$ we find $E \circ P$ is an elliptic transformation of order 3, which implies that $(EP)^3$ is the other relation. $\square$



The above figure demonstrates the side-pairing algorithm for detecting relations applied to the marked vertex v_1.

6 Fuchsian limit sets

The defining feature of a Fuchsian group is that it acts discontinuously on hyperbolic space. However, a Fuchsian group (usually) acts continuously on part of the boundary! The precise subset of the boundary where the action is continuous is known as the *limit set*. Fuchsian limit sets are interesting in their own right and often have a complicated fractal structure, but they are also intrinsically linked with the algebraic and geometric properties of the group and its action. It is more convenient to set up the theory for Fuchsian groups acting on $\mathbb{D}^2$, but of course one can pass between the models via the Cayley map as usual.

More precisely, the limit set of a Fuchsian group $\Gamma \leqslant \mathrm{con}^+(1)$ is defined by

$$L(\Gamma) = \overline{\Gamma(0)} \setminus \Gamma(0)$$

where the closure is taken in the *Euclidean* metric. In other words, the limit set is the set of accumulation points of the orbit of 0 under the action of the group. A more direct, but more laborious, way to write the limit set is

$$L(\Gamma) = \{z \in \mathbb{C} \setminus \Gamma(0) \ : \ \text{there exists a sequence } g_n \in \Gamma \text{ such that } |g_n(0) - z| \to 0\}.$$

In this section we study Fuchsian limit sets $L(\Gamma)$ in detail and begin by making some simple structural observations.

Theorem 6.1. *Fuchsian limit sets are closed subsets of the boundary of hyperbolic space, i.e. if* $\Gamma \leqslant \mathrm{con}^+(1)$ *is a Fuchsian group, then*

1. $L(\Gamma) \subseteq S^1$
2. $L(\Gamma)$ *is closed in the Euclidean metric.*

Proof. 1. follows immediately since $\Gamma(0)$ is a discrete subset of $\mathbb{D}^2$. In particular, $\Gamma(0)$ cannot accumulate anywhere except on the boundary. We will now prove *2.* by proving that if $z_n \in L(\Gamma)$ is a sequence of points such that $z_n \to z$, then $z \in L(\Gamma)$. For each n, let $g_{n,m} \in \Gamma$ be such that $g_{n,m}(0) \to z_n$ in the Euclidean metric as $m \to \infty$. We can find such $g_{n,m}$ since $z_n \in L(\Gamma)$. We may also assume that $g_{n,m}(0)$ converges 'uniformly quickly' to z_n by taking appropriate subsequences. More precisely, for a given n choose m_n sufficiently large to ensure that

$$|g_{n,m_n}(0) - z_n| \leqslant 1/n.$$

Then

$$|g_{n,m_n}(0) - z| \leqslant |g_{n,m_n}(0) - z_n| + |z_n - z| \leqslant 1/n + |z_n - z| \to 0$$

as $n \to \infty$ and therefore $z \in L(\Gamma)$. □

At this point it seems that the limit set may depend on the choice of 'base point'. The base point above is 0, but we could have defined the limit set to be the accumulation points of any $w \in \mathbb{D}^2$. It turns out that the limit set is independent of the base point and, moreover, the same sequence of maps will give rise to the same point in the limit set, independent of the base point.

Theorem 6.2. *Fuchsian limit sets are independent of the choice of base point, i.e. if* $\Gamma \leqslant \mathrm{con}^+(1)$ *is a Fuchsian group and* $w \in \mathbb{D}^2$*, then*

$$L(\Gamma) = \overline{\Gamma(w)} \setminus \Gamma(w).$$

Proof. Let $z \in L(\Gamma)$ and let $g_n \in \Gamma$ be such that $g_n(0) \to z$. We will prove that $g_n(w) \to z$ for an arbitrary point $w \in \mathbb{D}^2$, which proves that $L(\Gamma) \subseteq \overline{\Gamma(w)} \setminus \Gamma(w)$ as required. The opposite inclusion follows via a similar argument which we omit. Using $\mathrm{con}^+(1)$ invariance and one of the standard formulae for the distance between two points we know that

$$\frac{|g_n(w) - g_n(0)|^2}{(1 - |g_n(w)|^2)(1 - |g_n(0)|^2)} = \sinh^2 \frac{d_{\mathbb{D}^2}(g_n(w), g_n(0))}{2} = \sinh^2 \frac{d_{\mathbb{D}^2}(w, 0)}{2} = \frac{|w|^2}{1 - |w|^2}.$$

Since $g_n(0) \to z \in S^1$ we know $(1 - |g_n(0)|^2) \to 0$. This forces $|g_n(w) - g_n(0)| \to 0$ as $n \to \infty$ as otherwise the left hand side of the above equation would blow up. Therefore

$$|g_n(w) - z| \leqslant |g_n(w) - g_n(0)| + |g_n(0) - z| \to 0$$

as required. □

A common theme in fractal geometry is to use the fact that the fractal may be a 'dynamical invariant'. For example, the middle third Cantor set is invariant under the map $x \mapsto 3x$ (mod 1). One can then appeal to dynamical systems theory to study the fractal. Limit sets are invariant under the action of the associated Fuchsian group, and this property is often used to study the limit sets and the invariant measures they support.

Theorem 6.3. *Limit sets are (strongly) invariant under the action of the associated Fuchsian group, i.e. if* $\Gamma \leqslant \mathrm{con}^+(1)$ *is a Fuchsian group, then for all* $g \in \Gamma$ *we have*

$$g(L(\Gamma)) = L(\Gamma).$$

Proof. Let $z \in L(\Gamma)$, $g \in \Gamma$, and let $g_n \in \Gamma$ be such that $g_n(0) \to z$. Since g is a continuous map on the Euclidean disk, it follows that

$$g(g_n(0)) \to g(z)$$

which implies $g(z) \in L(\Gamma)$. Thus we have proved that $g(L(\Gamma)) \subseteq L(\Gamma)$. The opposite inclusion follows by replacing g with g^{-1} and applying g. □

After establishing Γ-invariance of the limit set, it should not come as a surprise that the limit set contains all the hyperbolic and parabolic fixed points of elements of Γ. Of course, it cannot contain the elliptic fixed points because they lie in the interior of the disk.

Theorem 6.4. *Limit sets contain all hyperbolic and parabolic fixed points, i.e. if* $\Gamma \leqslant \mathrm{con}^+(1)$ *is a Fuchsian group and* $z \in S^1$ *is fixed by some* $g \in \Gamma$*, then* $z \in L(\Gamma)$.

Proof. This follows immediately since if $p \in S^1$ is fixed by a parabolic map $g \in \Gamma$, then $g^n(0) \to p$ and if $h_1, h_2 \in S^1$ are the repelling and attracting fixed points of a hyperbolic map $g \in \Gamma$, then

$$g^n(0) \to h_2 \qquad \text{and} \qquad g^{-n}(0) \to h_1.$$

The precise calculation is in the tutorial sheets. □

Limit sets can be very complicated metric objects, but from a topological point of view there are only 5 possibilities. More precisely, only 5 different topological spaces can arise as limit sets of Fuchsian groups. The most interesting of these is the 'Cantor space'. A subset of a metric space is a *topological Cantor space* if it is compact, totally disconnected, and perfect. Recall that a subset of a metric space is *totally disconnected* if the only connected subsets are singletons. In particular, a subset of S^1 is totally disconnected if and only if given any two points in the set there is a point not in the set on both the clockwise and anti-clockwise arcs between the two given points. A subset of a metric space is *perfect* if every point is an accumulation point. An important result in topology is that being compact, totally disconnected, and perfect uniquely defines a topological space up to homeomorphism.

Theorem 6.5. *Let $\Gamma \leqslant \mathrm{con}^+(1)$ be a Fuchsian group. Then*

1. *$L(\Gamma) = \emptyset$ if and only if Γ consists only of elliptic elements,*
2. *$|L(\Gamma)| = 1$ if and only if Γ is generated by a single parabolic element,*
3. *$|L(\Gamma)| = 2$ if and only if Γ is generated by a single hyperbolic element, or if it is generated by a single hyperbolic element and an elliptic involution (element of order 2) which interchanges the hyperbolic fixed points.*
4. *in all other cases $L(\Gamma)$ is either equal to S^1 or is a topological Cantor space. In particular, $L(\Gamma)$ is a continuum (has the same cardinality as $\mathbb{R}$).*

Proof. If $L(\Gamma) = \emptyset$, then Γ has no fixed points on the boundary and therefore must consist only of elliptic elements.

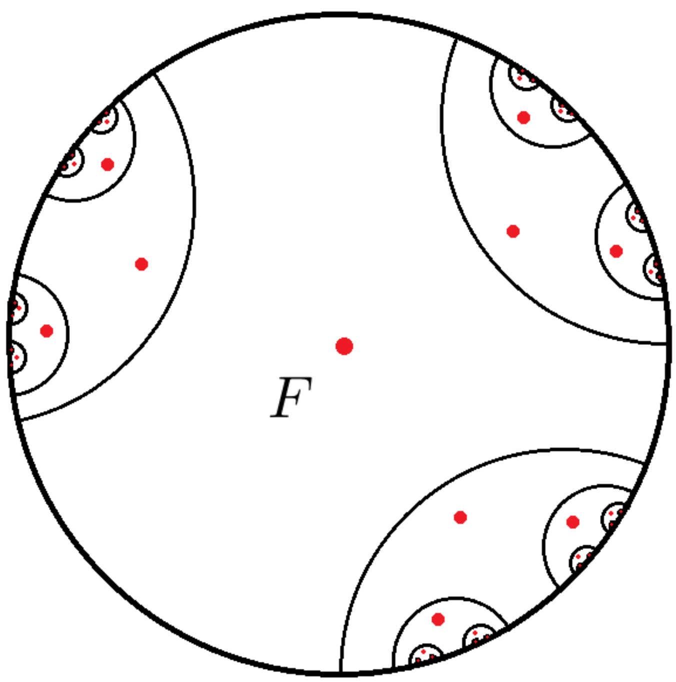
If $|L(\Gamma)| = 1$, then there are no hyperbolic elements since they have 2 fixed points. Also there must be at least one parabolic element and all parabolic elements must fix the same point. Since Γ is discrete, this guarantees that all parabolic elements are powers of a common parabolic element. Finally, there are no elliptic elements because a singleton cannot be invariant under a non-trivial rotation.

If $|L(\Gamma)| = 2$, then there cannot be any parabolic elements since they fix precisely one point. If there are no elliptic elements, then every element must fix the two points in the limit set and again using discreteness must all be powers of the same hyperbolic element. Moreover, elliptic elements which do exist must interchange the two fixed points and are thus involutions. Finally, there is a unique involution interchanging any two given points in S^1.

Let Γ be a Fuchsian group whose limit set has at least 3 points. Fix any three distinct points $z_1, z_2, z_3 \in L(\Gamma)$ and let $w \in \mathbb{D}^2$ be any point on the geodesic ray joining z_2 and z_3. Our aim is to show that z_1 is an accumulation point in $L(\Gamma)$. Let $g_n \in \Gamma$ be distinct elements such that $g_n(w) \to z_1$ (recall the limit set is independent of the base point). Take a subsequence of the g_n such that $g_n(z_2)$ and $g_n(z_3)$ both converge in the Euclidean metric to (not necessarily distinct) points $\tilde{z}_2, \tilde{z}_3 \in S^1$ respectively (we can do this by compactness). This implies that $g_n(w)$ converges in the Euclidean metric to a point on the geodesic ray joining $\tilde{z}_2$ and $\tilde{z}_3$. However, since we already know $g_n(w) \to z_1$ we may conclude that at least one of $\tilde{z}_2$ and $\tilde{z}_3$ is equal to z_1. We assume without loss of generality that $\tilde{z}_2 = z_1$. If at most finitely many n satisfy $g_n(z_2) = z_1$, then we can conclude that z_1 is an accumulation point. This follows since $\Gamma(z_2) \subseteq L(\Gamma)$ by Γ-invariance. Therefore, we assume without loss of generality that $g_n(z_2) = z_1$ for all n. Since elements of $\mathrm{con}^+(1)$ are determined by their action on two points, it follows that the points $g_n(z_3)$ are all distinct and therefore $\tilde{z}_3$ is an accumulation point in $L(\Gamma)$. If $\tilde{z}_3 = z_1$, then again we can conclude that z_1 is an accumulation point and so we further assume that $\tilde{z}_3 \neq z_1$. The strategy now is to drag $\tilde{z}_3$ towards z_1. Consider the (distinct) points z_3 and $g_1^{-1}(\tilde{z}_3)$. We know $g_n(z_3) \to \tilde{z}_3 \neq z_1$ and so $g_n(g_1^{-1}(\tilde{z}_3)) \to z_1$ by repeating the above argument. Since $\tilde{z}_3$ is an accumulation point, we know $g_1^{-1}(\tilde{z}_3)$ is an accumulation point and so z_1 is the limit of accumulation points and so is itself an accumulation point. Since z_1 was arbitrary we conclude that $L(\Gamma)$ is perfect.

We already know that $L(\Gamma)$ is closed and bounded, and thus compact, and so it remains to show that if $L(\Gamma)$ has at least three points and $L(\Gamma) \neq S^1$, then it is totally disconnected. Let $z_1, z_2 \in L(\Gamma)$ be distinct points and suppose $u \in S^1 \setminus L(\Gamma)$. It suffices to show that there is a point in the clockwise arc from z_1 to z_2 which is not in $L(\Gamma)$. The anti-clockwise arc can be dealt with similarly. Since $L(\Gamma)$ is closed, we can find an open neighbourhood of u which is disjoint from $L(\Gamma)$. Let u_1, u_2 be the end points of this open neighbourhood (which is an open arc in S^1). Let $t \in S^1$ be an arbitrary point in the clockwise arc from z_1 to z_2. If $t \notin L(\Gamma)$, then we are done and so we can assume $t \in L(\Gamma)$. Arguing as above we can find a sequence of maps $g_n \in \Gamma$ such that either $g_n(u_1)$ or $g_n(u_2)$ converges to t in the Euclidean metric. Using strong Γ-invariance again this guarantees the existence of a point in the clockwise arc from z_1 to z_2 which is not in $L(\Gamma)$. □

We say a Fuchsian group is an *elementary Fuchsian group* if it has a finite limit set (i.e. we are in one of the first three cases in the above theorem). Otherwise it is called a *non-elementary Fuchsian group*. Note that in both the second and third cases (when there are no elliptic elements, $\Gamma \cong (\mathbb{Z}, +)$, the infinite cyclic group. This is a simple demonstration that the limit set is not a purely group theoretic concept (invariant under group isomorphism).



The above figure depicts a subgroup of con(1) generated by reflection in three geodesic rays. Groups generated in this way are called (hyperbolic) reflection groups. The orbit of 0 is shown in red and the fundamental domain is labelled F. Note that reflection groups are strictly speaking not Fuchsian because reflections are orientation reversing. However, the orientation preserving elements of a reflection group form a Fuchsian subgroup of index 2 and the limit set is the same for both groups.

6.1 The Poincaré series and dimension of the limit set

The Poincaré series of a Fuchsian group $\Gamma \leqslant \mathrm{con}^+(1)$ measures how quickly the orbit $\Gamma(0)$ accumulates on the boundary. Heuristically, the faster the accumulation, the smaller the limit set is. More formally, we define a function $P : [0, \infty) \to [0, \infty]$ by

$$P_\Gamma(s) = \sum_{g \in \Gamma} e^{-s d_{\mathbb{D}^2}(0, g(0))} = \sum_{z \in \Gamma(0)} e^{-s d_{\mathbb{D}^2}(0, z)} = \sum_{z \in \Gamma(0)} \left(\frac{1 - |z|}{1 + |z|} \right)^s$$

and refer to $P_\Gamma(s)$ as the Poincaré series of Γ with exponent $s \geqslant 0$. The faster the orbit $\Gamma(z)$ accumulates at the boundary, the smaller $P(s)$ is and in order to quantify this precisely, the *exponent of convergence* of Γ is defined by

$$\delta(\Gamma) = \inf\{s \geqslant 0 \ : \ P_\Gamma(s) < \infty\}.$$

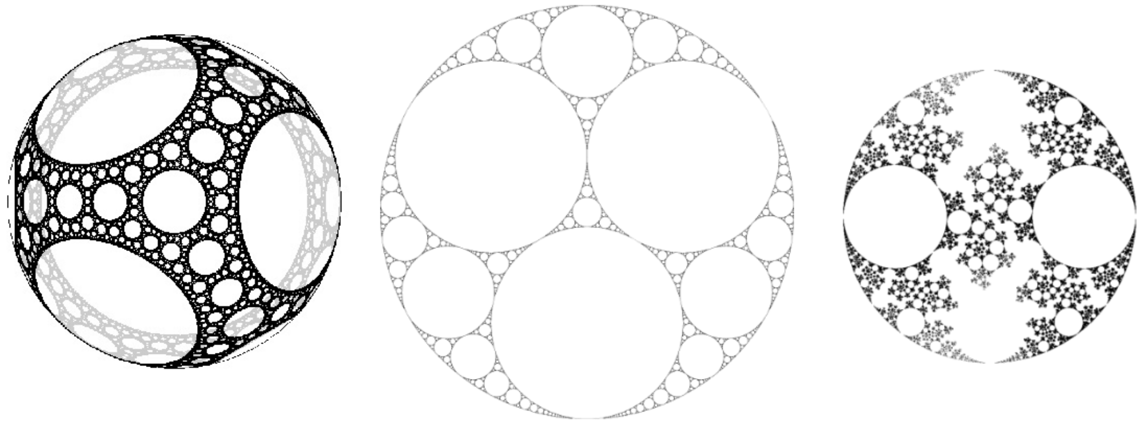
The most standard way to quantify the 'size' of a fractal set is by its 'dimension'. The Hausdorff dimension is defined for any subset of a metric space and generalises our intuitive conception of the dimension of smooth objects to fractals. For example, the Hausdorff dimension of any countable set is 0, the Hausdorff dimension of S^1 is 1, the Hausdorff dimension of the Euclidean plane $\mathbb{C}$ is 2, but the Hausdorff dimension of the (non-smooth) middle third Cantor set is $\log 2 / \log 3 \approx 0.6309\ldots$. Unfortunately, the precise definition and study of the Hausdorff dimension is beyond the scope of this course, but we end with a beautiful (and comparatively recent) result concerning the challenging question of computing the dimension of a Fuchsian limit set. We write $\dim_{\mathrm{H}} E$ to denote the Hausdorff dimension of a set E.

Theorem 6.6. *If Γ is a finitely generated non-elementary Fuchsian group, then*

$$\dim_{\mathrm{H}} L(\Gamma) = \delta(\Gamma).$$

This important theorem dates back to an influential paper of Samuel J. Patterson from 1976 (*Acta Mathematica*). Patterson received his PhD from Cambridge in 1975 under the supervision of Alan Beardon (author of one of the main references for this course) and has worked in Göttingen since 1981. It was here that he supervised the PhD of Bernd Stratmann, who later joined the University of St Andrews and taught me hyperbolic geometry in 2009!

This theorem holds for certain classes of infinitely generated Fuchsian groups, including *geometrically finite* groups. Roughly speaking a Fuchsian group is geometrically finite if it has a 'reasonable' fundamental domain. The result was generalised to higher dimensional hyperbolic spaces (where the analogous groups are called *Kleinian groups*) by Dennis Sullivan in 1979. Those familiar with fractal geometry and dimension theory will know that there are several other notions of dimension apart from the Hausdorff dimension. In 1996 Stratmann and Urbański proved that the upper and lower box-counting dimensions and packing dimension of the limit set of a geometrically finite non-elementary Kleinian group is also equal to the exponent of convergence (and thus coincide with the Hausdorff dimension).



The above figure shows three limit sets of Kleinian groups acting on three dimensional hyperbolic space. The one on the left is a discrete subgroup of $\text{con}^+(2)$ which acts on $\mathbb{D}^3$ (the three dimensional disk). The boundary of this space is S^2 and the limit set is a fractal subset of the boundary. The other two images are subsets of $\mathbb{C}$, which is the boundary of three dimensional hyperbolic space when modelled by

$$\mathbb{H}^3 = \mathbb{R}^+ \times \mathbb{C}$$

(the three dimensional analogue of the upper half-plane model). The group of (orientation preserving) isometries in this case is $\text{PSL}(2, \mathbb{C})$. The image in the middle is more commonly recognised as an *Apollonian circle packing*, constructed as follows: begin with four mutually tangent circles, one of which contains the other three. Now draw the biggest circle you can inside the large initial circle, whose interior does not overlap the interiors of the three circles initially inside the large one. This circle will necessarily be tangent to three of the four initial circles. Continue in this way *ad infinitum* and the set of all the boundaries of circles you have added (including the initial four) is the Apollonian circle packing. Although the precise packing you obtain depends on the location of the initial four circles, it turns out that any Apollonian packing can be mapped onto any other via a Möbius map (recall that circles are mapped to circles). See Mark Pollicott's expository paper, found here `http://homepages.warwick.ac.uk/~masdbl/apollo-29Dec2014.pdf`, for more details on the history and persistent mathematical interest in Apollonian circle packings.