The Arithmetic of Kleinian Groups

April 20, 2016

Contents

1 Overview

1.1 Course Description

A *Kleinian group* is a discrete subgroup of $\mathrm{PSL}_2(\mathbb{C})$. Examples include

1. $\mathrm{PSL}_2(\mathbb{Z})$,
2. the Apollonian group, whose limit set is shown here:
3. the Bianchi groups $\mathrm{PSL}_2(\mathcal{O}_K)$ where $\mathcal{O}_K$ is the ring of integers of an imaginary quadratic field K.

There are arithmetic invariants associated to a Kleinian group, specifically, the invariant trace field and invariant quaternion algebra. The purpose of the course is to the study the relationship between these arithmetic invariants and the geometry of Kleinian groups. In order to study this, we will first need to study quaternion algebras, Möbius transformations, and hyperbolic geometry.

1.2 Pre-requisites

I will assume a solid working knowledge of algebraic number theory, e.g. as in Samuel's *Algebraic Theory of Numbers* and Chapter 0 of Maclachlan and Reid, including basic local theory. I will also assume knowledge of other graduate pillar topics (topology, analysis) as needed.

1.3 Resources

1. Maclachlan and Reid, *The arithmetic of hyperbolic 3-manifolds*. This is the main resource for the latter portion of the course. Chapter 0 will be assumed.

2. Beardon, *The geometry of discrete groups.* This is the main resource for the background in hyperbolic geometry and Kleinian groups in the first portion of the course. Chapter 1 will be assumed.

3. Samuel, *Algebraic theory of numbers.* This covers the necessary algebraic number theory background. It will be assumed.

4. Marden, *Outer circles: an introduction to hyperbolic* 3*-manifolds.* This introduces Möbius transformations and Kleinian groups and their relationship to hyperbolic 3-manifolds. It also has lots of 'explorations' that are suitable for further work for credit. It is good for an overview, but lacks details.

5. Parker, *Hyperbolic Spaces: The Jyväskylä Notes.* Available at `maths.dur.ac.uk/~dma0jrp/img/HSjyvaskyla.pdf`. This introduces the hyperbolic upper half plane and space in a nice algebraic way.

6. Katok, *Fuchsian Groups.* This does only the Fuchsian (H^2) case, but it is pleasantly compact.

1.4 Credit

Students wishing to receive full credit (an A) for the course shall

1. attend lecture regularly, and

2. demonstrate engagement with the material. Demonstrating engagement will be in the form of one of the following:

   (a) Demonstrating evidence of significant homework exercises (suggested exercises will be provided), numbering at least 15. These needn't be distributed throughout the full course, allowing students to concentrate on one topic (e.g. Möbius transformations or quaternion algebras) to mastery, if desired. However, students should respect their own time sufficiently so that they choose exercises they will learn something useful from.

   (b) Learning and presenting, in class, a further topic or related material. For example, one of the examples or applications of Chapters 4 and 5 of Maclachlan and Reid. This will involve producing written notes to be handed out to the class. It should involve at least one lecture, but more is fine if appropriate to the topic.

   (c) Covering for me while I'm gone at a conference; this will involve digesting and presenting course notes I leave for you. This should involve at least two lectures.

   (d) Any other agreed-upon demonstration.

Students taking the course for credit/no credit are also expected to attend regularly and do something to demonstrate engagement, but it can be smaller in scale.

A note on exercises: the ones suggested in the text vary greatly in difficulty, length and scope. Accordingly, the number 15 is only a guide. You may also choose exercises from the books, or make up your own.

2 Matrix groups, General and Special Linear Groups

2.1 Topological groups

A group G which is also a topological space, and such that its multiplication $G \times G \to G, (x, y) \mapsto xy$ and its inverse $G \to G, x \mapsto x^{-1}$ are continuous maps, is called a *topological group.*

Exercise 1. *A subgroup of a topological group is a topological group under the subspace topology.*

Exercise 2. *Let $g \in G$, where G is a topological group. Multiplication by g is continuous.*

Exercise 3. *Give an example of a group G, and a topology on it which makes multiplication continuous but under which the inverse fails to be continuous.*

2.2 Linear Groups

Given an integral domain R (generally, a subring of $\mathbb{C}$), we define

$$M_2(R) = \left\{ \begin{pmatrix} a & c \\ b & d \end{pmatrix}, a, b, c, d \in R \right\},$$

$$\mathrm{GL}_2(R) = \left\{ \begin{pmatrix} a & c \\ b & d \end{pmatrix}, a, b, c, d \in R, ad - bc \in R^* \right\},$$

and also

$$\mathrm{SL}_2(R) = \left\{ \begin{pmatrix} a & c \\ b & d \end{pmatrix}, a, b, c, d \in R, ad - bc = 1 \right\}.$$

Importantly, for $R = \mathbb{C}$, these last two are topological groups under matrix multiplication and an appropriate topology. We will define a metric on $M_2(\mathbb{C})$, given by the bilinear product of the trace pairing. Let A^* denote the conjugate transpose of A, i.e. $\overline{A}^t$.

Definition 4. *Let $A, B \in M_2(\mathbb{C})$. Define*

$$[A, B] = tr(AB^*),$$

called the trace pairing.

Proposition 5. *The trace pairing is a Hermitian form on the complex vector space $M_2(\mathbb{C})$. Furthermore, it induces a norm $||A|| = [A, A]^{1/2}$, which in turn induces a metric on $M_2(\mathbb{C})$, and hence a topology.*

Furthermore, we have some relationships with the metric on $\mathbb{C}$ and matrix multiplication:

1. $|\det A| \cdot ||A^{-1}|| = ||A||$
2. $||[A, B]|| \leq ||A|| \cdot ||B||$
3. $||AB|| \leq ||A|| \cdot ||B||$
4. $2|\det A| \leq ||A||^2$

Recall that a Hermitian form is a pairing which is linear in the first variable and satisfies $\overline{[A, B]} = [B, A]$, so that it is conjugate-linear in the second variable.

The norm just defined is often called the *Frobenius norm.* It is just the usual L^2 norm if we identify $M_2(\mathbb{C})$ with $\mathbb{C}^4$ by listing the matrix entries in a column vector. We will call the induced metric the *Frobenius norm metric* or more simply the *Frobenius metric.*

Proof. The proof is computational; see Beardon, §2.2 $\square$

The metric has the property that convergence is entrywise in the entries of the matrix.

Proposition 6. *Determinant and trace are continuous functions.*

Also, a norm is always continuous for a metric it defines.

Proposition 7. *The group $\mathrm{GL}_2(\mathbb{C})$ is a topological group under the Frobenius metric.*

Proof. It suffices to show that the maps $A \mapsto A^{-1}$ and $(A, B) \mapsto AB$ are continuous. Convergence in $\mathrm{GL}_2(\mathbb{C})$ is entrywise, and continuity is equivalently defined (for a metric space) by the preservation of limits. This reduces the question to one of continuity of the formulae for the entries. $\square$

Proposition 8. *The group $\mathrm{SL}_2(\mathbb{C})$ is a topological group under the Frobenius metric.*

Proof. The group $\mathrm{SL}_2(\mathbb{C})$ is clearly a subgroup of $\mathrm{GL}_2(\mathbb{C})$ and therefore is a topological group under the subspace topology, which is again given by the Frobenius metric. $\square$

Exercise 9. *Which, if any, of the theory above relies specifically on the dimension 2×2 of concern to us? Does it all work for $n \times n$?*

2.3 Discrete groups

Definition 10. *A topological group is called* discrete *if the topology is the discrete topology.*

In particular, a subgroup H of a topological group G is discrete if the subspace topology induced on H is the discrete topology.

cretestabilizes

Proposition 11. *A subgroup H of G is discrete if and only if any sequence in H converging to some $g \in G$ eventually stabilizes.*

Proof. Suppose g_n is a such a sequence. Then $g_n \to g \in G$. Then $g_n g_{n+1}^{-1} \to gg^{-1} = I$. If H is discrete, then the identity is isolated, and the sequence $g_n g_{n+1}^{-1}$ is eventually equal to I, so that g_n is eventually constant.

If H is not discrete, then the identity is not isolated, and there exists a sequence $a_n \to I$ within H, where $a_n \neq I$ for all n, so that this sequence does not eventually stabilize. □

Exercise 12. *Is it more generally true that a topological group with a metric topology is discrete if and only if any Cauchy sequence of elements eventually stabilizes?*

Proposition 13. *Let G be a topological group with an open singleton. Then every singleton is open and G is discrete.*

Proof. This is an immediate consequence of the fact that if U is open, then gU is open (which follows since gU is the preimage of U under multiplication by g^{-1}). □

Exercise 14. *It is a more general fact that multiplication-by-g is a homeomorphism on a topological group. Prove it.*

By the previous proposition, it suffices to verify that one point of H, for example the identity, is isolated to conclude that H is a discrete subgroup.

A typical example is that the *Modular group*,

$$\mathrm{SL}_2(\mathbb{Z}) = \left\{ \begin{pmatrix} a & c \\ b & d \end{pmatrix} : a, b, c, d \in \mathbb{Z}, ad - bc = 1 \right\}$$

is a discrete subgroup of $\mathrm{SL}_2(\mathbb{C})$. To verify that it is discrete, we can check that the identity is isolated, i.e.

$$\inf\{||X - I|| : X \neq I, X \in \mathrm{SL}_2(\mathbb{Z})\} \geq \inf\{||X|| : X \neq 0, X \in M_2(\mathbb{Z})\} = 1.$$

Proposition 15. *Let K be a number field with ring of integers $\mathcal{O}_K$, embedded as a subring of $\mathbb{C}$. Then $\mathrm{GL}_2(\mathcal{O}_K) < \mathrm{GL}_2(\mathbb{C})$ is discrete if and only if K is $\mathbb{Q}$ or an imaginary quadratic field. The same is true for $\mathrm{SL}_2(\mathcal{O}_K) < \mathrm{GL}_2(\mathbb{C})$.*

Proof. Consider $\mathcal{O}_K$ as a subgroup of $\mathbb{C}$ (as an additive topological group under the usual metric on $\mathbb{C}$). If K is $\mathbb{Q}$ or imaginary quadratic, then $\mathcal{O}_K$ is discrete. Therefore $\inf\{||X|| : X \neq 0, X \in M_2(\mathcal{O}_K)\}$ is bounded below, so the previous argument applies.

Now conversely, if K is neither $\mathbb{Q}$ nor imaginary quadratic, then $\mathcal{O}_K$ has positive unit rank. Take a unit u of infinite order. If $|u| = 1$, then since the unit circle is compact, there is a convergent subsequence of u^n (which doesn't stabilize). Otherwise, powers of u (or u^{-1}) will approach 0. In either case, we have a_n a sequence in $\mathcal{O}_K^*$ converging in $\mathbb{C}$ which does not stabilize.

Consider the sequence

$$\begin{pmatrix} 1 & a_n \\ 0 & 1 \end{pmatrix} \in \mathrm{SL}_2(\mathcal{O}_K).$$

This sequence does not stabilize, but converges in $\mathrm{GL}_2(\mathbb{C})$. Therefore, $\mathrm{SL}_2(\mathcal{O}_K)$, and hence $\mathrm{GL}_2(\mathcal{O}_K)$, are not discrete. □

This proof also demonstrates that discreteness of $\mathcal{O}_K$ and $\mathrm{GL}_2(\mathcal{O}_K)$ are independent of the embedding of $\mathcal{O}_K$.

:discreteheight **Proposition 16.** *Let H be a subgroup of $\mathrm{SL}_2(\mathbb{C})$. Then H is discrete if and only if for each positive k, $\{X \in H : ||X|| \leq k\}$ is finite.*

Proof. Suppose that $A_k := \{x \in H : ||x|| \leq k\}$ is finite for all k. Let X have norm ℓ for some ℓ. Then X is bounded away from all other points in the finite set $A_{\lceil \ell \rceil + 1}$, and from the complement of $A_{\lceil \ell \rceil + 1}$. Therefore we have isolated a point of H, showing H is discrete. (Alternatively, the norm function is continuous, so any sequence of elements x_n in H which is not eventually constant but has a limit $x_n \to g$ would result in an infinite A_k, where $k > ||g||$.)

Conversely, suppose A_k is infinite for some k. Then, it contains an infinite sequence of distinct elements of H; as the ball of radius k is compact, this sequence contains a convergent subsequence inside $M_2(\mathbb{C})$; the limit must lie in $\mathrm{SL}_2(\mathbb{C})$ since the determinant is continuous. But this sequence is, by assumption, not eventually constant. So H must not be discrete, by Proposition 11. □

In particular, any discrete subgroup of $\mathrm{SL}_2(\mathbb{C})$, as a union of the finite sets A_k, is countable.

Exercise 17. *The proof above uses the fact that we are dealing with $\mathrm{SL}_2(\mathbb{C})$; give a counterexample when we replace it with $\mathrm{GL}_2(\mathbb{C})$. (For a hint, see Beardon, Exercise 2.3.1.)*

Exercise 18. *If H is a discrete subgroup of G, and $g \in G$, then gHg^{-1} is discrete.*

Exercise 19. *Any of Beardon, Exercises of §2.3.*

3 Complex Möbius transformations

We've now examined the group $\mathrm{GL}_2(\mathbb{C})$ in some detail, discussing not just its group structure but its topological group structure and its discrete subgroups. But we should ask, what is $\mathrm{GL}_2(\mathbb{C})$ acting on? Most obviously, it acts on $\mathbb{C}^2$. But we will be concerned instead with a quotient of $\mathbb{C}^2$, called $\mathbb{P}^1(\mathbb{C})$.

3.1 $\mathbb{P}^1(\mathbb{C})$ and the chordal metric

We may define $\mathbb{P}^1(\mathbb{C})$ to be $(\mathbb{C}^2 \smallsetminus \{(0,0)\})/\sim$ where $(X,Y) \sim (Z,W)$ if there exists some $u \in \mathbb{C}^*$ so that $X = uZ$ and $Y = uW$. Equivalently, $XW = ZY$. This is the space of lines of $\mathbb{C}^2$ (i.e. one-dimensional subspaces in the sense of complex vector spaces).

We can identify $\mathbb{P}^1(\mathbb{C})$ with $\widehat{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$ via the map

$$(X,Y) \mapsto X/Y.$$

This has inverse $z \mapsto [z,1]$ on $\mathbb{C}$ and $\infty \mapsto [1,0]$.

We may also identify $\widehat{\mathbb{C}}$ with the so-called *Riemann sphere*, the unit sphere in $\mathbb{R}^3$, writing $z \mapsto z^*$. To do this, identify $\mathbb{C}$ with the xy plane in $\mathbb{R}^3$ (in coordinates x, y, t) via $z = x + iy \mapsto (x, y, 0)$. Then we map the xy plane to $S^2 \smallsetminus (0,0,1)$ as follows. The point z^* is defined to be the intersection of the line from $(0,0,1)$ to $(x,y,0)$ with S^2. Finally, we identify $(0,0,1)$ with ∞. This map $S^2 \to \widehat{\mathbb{C}}$ and its inverse are both (!) known as *stereographic projection*. The formula for stereographic projection $\widehat{\mathbb{C}} \to S^2$ is

$$z = x + iy \mapsto \left(\frac{2x}{|z|^2+1}, \frac{2y}{|z|^2+1}, \frac{|z|^2-1}{|z|^2+1} \right),$$

and its inverse is

$$(a,b,c) \mapsto \frac{a}{1-c} + \frac{b}{1-c} i.$$

Exercise 20. *Verify the formula. To complete the triangle, of course we can map $\mathbb{P}^1(\mathbb{C})$ to S^2 and vice versa. Do this explicitly.*

One can also consider $\widehat{\mathbb{C}}$ to be a complex manifold, which is also referred to as the Riemann sphere.

If one considers $\mathbb{C}^2$ as a real vector space, and identifies vectors of $\mathbb{C}^2 \setminus \{(0,0)\}$ which are real multiples (i.e. take the collection of one-dimensional real subspaces), then one obtains S^3. Further identifying complex multiples, one is left with S^2. This map $S^3 \to S^2$ has kernel S^1 and is called the *Hopf fibration*: you can watch an amazing video about it at this url: http://nilesjohnson.net/hopf.html.

The points on S^2 inherit the usual metric from the ambient space $\mathbb{R}^3$. The induced metric on $\widehat{\mathbb{C}}$ is called the *chordal metric*. More precisely, for $z, w \in \widehat{\mathbb{C}}$,

$$\rho(z,w) = \frac{1}{2} \left| z^* - w^* \right|.$$

Exercise 21. *This works out to be, for $z, w \neq \infty$,*

$$\rho(z,w) = \frac{|z-w|}{\sqrt{|z|^2+1}\sqrt{|w|^2+1}}.$$

In homogeneous coordinates,

$$\rho([z_1,z_2],[w_1,w_2]) = \frac{|z_1w_2 - w_1z_2|}{\sqrt{|z_1|^2+|z_2|^2}\sqrt{|w_1|^2+|w_2|^2}}.$$

The point of this metric is to give a topology to $\widehat{\mathbb{C}}$, and in particular one in which ∞ is not distinguished. But it will also occasionally be useful as a metric.

Exercise 22. *Restricted to $\mathbb{C}$, the chordal topology is the usual topology. It is also the topology one obtains from $\mathbb{P}^1(\mathbb{C})$, which has a topology as a quotient of $\mathbb{C}^2 \setminus \{(0,0)\}$.*

Later, we will see that Möbius transformations are homeomorphisms with respect to this topology.

We can extend the maps of addition and multiplication on $\mathbb{C}$ to $\widehat{\mathbb{C}}$ by working homogeneously:

$$\begin{aligned}
[X,Y]+[A,B] &= [XB+AY, YB] \\
[X,Y]-[A,B] &= [XB-AY, YB] \\
[X,Y]\cdot[A,B] &= [XA, YB] \\
[X,Y]/[A,B] &= [XB, YA]
\end{aligned}$$

Of course, the result is no longer a ring. In particular, $1/0 = \infty$ and $1/\infty = 0$, while $\infty + z = \infty$ and so forth.

3.2 Circles and circle inversions in $\widehat{\mathbb{C}}$

We now wish to define a notion of *circle* in $\widehat{\mathbb{C}}$. My preferred definition is as follows.

def:circle **Definition 23.** *A* circle *is any locus of $\mathbb{P}^1(\mathbb{C})$ of the form*

$$\{[X,Y] : aX\overline{X} + bX\overline{Y} + \overline{b}\overline{X}Y + cY\overline{Y} = 0\}$$

where $b \in \mathbb{C}$, $a, c \in \mathbb{R}$, $ac - b\overline{b} < 0$.

Exercise 24. *What happens if $ac = b\overline{b}$ or $ac - b\overline{b} > 0$?*

As it turns out, considered in S^2, circles are just the intersections of S^2 with orthogonal spheres in the ambient $\mathbb{R}^3$.

When we wish to discuss these as living in $\widehat{\mathbb{C}}$, we notice a distinction: circles passing through ∞ (i.e. $a = 0$ in the definition above) become what we normally would call lines in $\mathbb{C}$, using the Euclidean geometry of $\mathbb{C}$; while the others become circles in the normal Euclidean sense in $\mathbb{C}$.

In other words, circles of $\widehat{\mathbb{C}}$ are either 'true circles' of the form

$$\begin{aligned} C(a,r) &= \{x \in \mathbb{C} : |x-a| = r\} \\ &= \{[X,Y] \in \mathbb{P}^1(\mathbb{C}) : X\overline{X} - a\overline{X}Y - \overline{a}X\overline{Y} + (a\overline{a} - r^2)Y\overline{Y} = 0\} \end{aligned}$$

where $r \neq 0$, or 'lines' of the form

$$\begin{aligned} L(a,r) &= \{x \in \mathbb{C} : \Re(x\overline{a}) = r\} \cup \{\infty\} \\ &= \{[X,Y] \in \mathbb{P}^1(\mathbb{C}) : \frac{\overline{a}}{2}X\overline{Y} + \frac{a}{2}\overline{X}Y - rY\overline{Y} = 0\} \end{aligned}$$

where $a \neq 0$. Note that, although I have reused the variables, a and r do not play exactly analogous roles in $C(a,r)$ and $L(a,r)$. (We write $\Re$ for real part and $\Im$ for imaginary part.)

These are *not* circles in the sense of the chordal metric.

Exercise 25. *Prove my claims above that true circles and lines in $\widehat{\mathbb{C}}$ correspond to intersections with spheres in S^2, and with the homogenous form given.*

Exercise 26. *What is the centre of a line?*

A *circle inversion* in plane geometry is a map fixing a circle of radius r, swapping its centre a with ∞, and taking any point p (besides the center a) to a point q on the same ray emanating from the centre so that

$$|p-a||q-a| = r^2$$

It is not hard to verify that inversion in the unit circle $C(0,1)$ in $\widehat{\mathbb{C}}$ is accomplished by

$$z \mapsto \frac{1}{\overline{z}} = \frac{z}{|z|},$$

and inversion in $C(0,r)$, $r \neq 0$, is accomplished by

$$z \mapsto \frac{r^2}{\overline{z}}.$$

Therefore, we obtain a more general inversion in $C(a,r)$ by

$$z \mapsto a + \frac{r^2}{\overline{\overline{z-a}}}.$$

Of course, these formulas are meaningful on all of $\widehat{\mathbb{C}}$ using the homogeneous coordinates – I won't mention this from now on.

One can also reflect in a line $L(a,r)$ where $a \neq 0$; the explicit formula is

$$z \mapsto z - 2\frac{\Re(z\overline{a}) - r}{\overline{a}}.$$

I will refer to this line reflection as a type of circle inversion, since lines are a type of circle.

Note that all of these are all of the form

$$z \mapsto \frac{a\overline{z} + b}{c\overline{z} - \overline{a}},$$

where $a \in \mathbb{C}$, $b, c \in \mathbb{R}$, $a\overline{a} + bc \neq 0$.

Next, observe that the fixed points of such a map form a circle. That is, the equality

$$[z, w] = [a\overline{z} + b\overline{w}, c\overline{z} - \overline{aw}]$$

in $\mathbb{P}^1(\mathbb{C})$ is exactly the equation of a circle, i.e.

$$c\overline{z}z - \overline{aw}z = a\overline{z}w + b\overline{w}w.$$

These observations motivate the following definition.

Definition 27. *A* circle inversion *on $\widehat{\mathbb{C}}$ is exactly the collection of homeomorphisms which can be expressed in the form*

$$z \mapsto \frac{a\overline{z} + b}{c\overline{z} - \overline{a}},$$

where $a \in \mathbb{C}$, $b, c \in \mathbb{R}$, $a\overline{a} + bc \neq 0$.

The inversions we've seen above all have this form. If $c \neq 0$, then dividing through top and bottom by c, we obtain the form of a plane circle inversion. If $c = 0$, then we obtain the given form of reflection in a line.

Proposition 28. *Circle inversions are homeomorphisms of $\widehat{\mathbb{C}}$.*

Proof. One can homogenize. That is,

$$z \mapsto \frac{a\overline{z} + b}{c\overline{z} + d}$$

should be written

$$[z, w] \mapsto [a\overline{z} + b\overline{w}, c\overline{z} + d\overline{w}],$$

which is clearly continuous. Checking continuity suffices, since these are inversions. Alternatively, one can check the limits at ∞ and $\phi(\infty)$ under the chordal metric as in Beardon, p.22. $\square$

For practice, let us now verify that, indeed, a reflection in a line becomes an inversion in a circle if one dehomogenizes at a different point. If 0 is not on the line, then the map

$$[z,w] \mapsto [a\overline{z} + b\overline{w}, c\overline{z} - \overline{aw}],$$

dehomogenized so that $[0,1]$ is the point at ∞, becomes

$$[1,w] \mapsto [a + b\overline{w}, c - \overline{aw}]$$

or

$$w \mapsto \frac{-\overline{aw} + c}{b\overline{w} + a},$$

which is again a circle inversion. You can dehomogenize elsewhere too if needed, for example, by applying a change of variables

$$[z,w] \mapsto [Z,W] = [z-w, w]$$

and then dehomogenizing at $[Z,W] = [0,1]$, i.e. $[z,w] = [1,1]$. This takes 1 as the point at ∞.

3.3 Möbius transformations

One can define Möbius transformations on the Riemann sphere, or, more generally, on any sphere S^n. That is, they are the group of maps $S^n \to S^n$ generated by reflections in $(n-1)$-dimensional spheres. This corresponds, via stereographic projection, to the group generated by inversions in spheres of $\widehat{\mathbb{R}^n}$. I will not do the most general theory, as Beardon does, but instead focus on the case of $n = 2$, and identify $\widehat{\mathbb{R}^2}$ with $\widehat{\mathbb{C}}$, or $\mathbb{P}^1(\mathbb{C})$.

Definition 29. *The* General Möbius group *on* $\widehat{\mathbb{C}}$ *is defined as the group of homeomorphisms generated by reflections. It is denoted* $\mathrm{GM}(\widehat{\mathbb{C}})$*.*

The Möbius group *on* $\widehat{\mathbb{C}}$ *is defined as the subgroup of the General Möbius group made up of orientation-preserving transformations. It is denoted* $\mathrm{M}(\widehat{\mathbb{C}})$*.*

Proposition 30. *Elements of the general Möbius group are all of the form*

$$z \mapsto \frac{a\overline{z} + b}{c\overline{z} + d},$$

or

$$z \mapsto \frac{az + b}{cz + d},$$

where $a, b, c, d \in \mathbb{C}$*,* $ad - bc \neq 0$*.*

Proof. We have verified that circle inversions are of this form, and it is an easy – but important – computation, that composition of things of this form is again of this form. For example, write

$$\gamma : z \mapsto \frac{a\overline{z} + b}{c\overline{z} + d},$$

and

$$\alpha : z \mapsto \frac{e\overline{z} + f}{g\overline{z} + h},$$

Then $\alpha \circ \gamma$ is

$$z \mapsto \frac{(e\overline{a} + f\overline{c})z + e\overline{b} + f\overline{d}}{(g\overline{a} + h\overline{c})z + g\overline{b} + h\overline{d}}.$$

Observe that the coefficients of $\alpha \circ \gamma$ are taken from the matrix

$$\begin{pmatrix} e & f \\ g & h \end{pmatrix} \begin{pmatrix} \overline{a} & \overline{b} \\ \overline{c} & \overline{d} \end{pmatrix}$$

and so the determinant is non-zero. Other cases are similar. □

The most important case of the computation above is that, if one composes the maps

$$\gamma : z \mapsto \frac{az + b}{cz + d},$$

and

$$\alpha : z \mapsto \frac{ez + f}{gz + h},$$

then $\alpha \circ \gamma$ is

$$z \mapsto \frac{(ea + fc)z + eb + fd}{(ga + hc)z + gb + hd}.$$

In other words, composing maps corresponds to matrix multiplication, since

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} e & f \\ g & h \end{pmatrix} = \begin{pmatrix} ea + fc & eb + fd \\ ga + hc & gb + hd \end{pmatrix}.$$

We will give names to some more types of transformations. A *scaling* is something of the form

$$z \mapsto az, \quad a \in \mathbb{C}^*.$$

A *translation* is of the form

$$z \mapsto z + a, \quad a \in \mathbb{C}.$$

Complex inversion is the map

$$z \mapsto \frac{1}{z}.$$

Proposition 31. *Scaling, translation and complex inversion are each compositions of an even number of circle inversions.*

Proof. Translation is accomplished by two reflections in lines prependicular to the direction of travel.

Scaling by a complex number is a composition of scaling by a real number and rotation about the origin. Scaling by a real number is a composition of inversion in two circles centred on the origin. Rotation is a composition of reflection in two lines passing through the origin.

And finally, complex inversion is a composition of inversion in the unit circle and complex conjugation. □

To get a visual sense of Möbius transformations, check out the video available at http://www.ima.umn.edu/~arnold/moebius/; you will also find a link to an article here, which explains their (alternate) definition of Möbius transformations.

Proposition 32. *Every transformation of the form*

$$z \mapsto \frac{az+b}{cz+d}, \quad ad-bc \neq 0, \quad a,b,c,d \in \mathbb{C}$$

is a composition of scalings, translations, and complex inversion. In particular, it is a composition of an even number of circle inversions, and therefore a Möbius transformation.

Proof. This is a computation. It so happens that, if $c \neq 0$, then

$$\frac{az+b}{cz+d} = \frac{a}{c} - \frac{ad-bc}{c} \cdot \frac{1}{cz+d}.$$

If $c = 0$, then since $ad-bc \neq 0$, it must be that $d \neq 0$. Then

$$\frac{az+b}{cz+d} = \frac{az+b}{d} = \frac{a}{d}z + \frac{b}{d}.$$

□

If $ad - bc = 0$, then the map becomes a constant.

Exercise 33. *This expresses the map in terms of at most 14 circle inversions; can you do fewer?*

Theorem 34. *The general Möbius group is*

$$\left\{ z \mapsto \frac{az+b}{cz+d} : \begin{array}{c} a,b,c,d \in \mathbb{C} \\ ad-bc \neq 0 \end{array} \right\} \cup \left\{ z \mapsto \frac{a\bar{z}+b}{c\bar{z}+d} : \begin{array}{c} a,b,c,d \in \mathbb{C} \\ ad-bc \neq 0 \end{array} \right\}.$$

Proof. As a transformation of the form

$$z \mapsto \frac{a\overline{z}+b}{c\overline{z}+d}$$

is a composition of one of the form

$$z \mapsto \frac{az+b}{cz+d}$$

together with complex conjugation, such transformations are in the general Möbius group. We have already verified that all circle inversions and their compositions are of one of these two forms. This determines the general Möbius group. □

3.4 Conformality and orientation

Definition 35. *A differentiable map $f : V \to V$, where V is a real or complex vector space, is called* conformal *if it preserves angles. One way to make this precise is to say that its Jacobian matrix (considering V as a real vector space) is a scalar multiple of an orthogonal matrix. We will say that f is* orientation preserving *if it preserves orientation, i.e. the Jacobian has positive determinant. Otherwise it is orientation reversing.*

As a brief reminder, a real $n \times n$ matrix is *orthogonal* if $|Ax| = |x|$ for all $x \in \mathbb{R}^n$, which occurs if and only if $A^{-1} = A^t$. A complex $n \times n$ matrix is *unitary* if $|Ax| = |x|$ for all $x \in \mathbb{C}^n$, which occurs if and only if $A^{-1} = \overline{A}^t$.

Therefore the idea behind the definition is that the best approximation to $f(x)$ near p is $f(p) + Jac(p)(x - p)$, i.e. an orthogonal transformation followed by dilation.

Note that the product of two orientation reversing transformations is orientation preserving.

If $V = \mathbb{C}$, a holomorphic function is conformal and orientation preserving away from the places where its derivative vanishes (this is an immediate consequence of the Cauchy-Riemann equations).

On the other hand, $f(z) = \overline{z}$ is not holomorphic but is conformal, since the map $x + iy \mapsto x - iy$ has Jacobian matrix

$$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

as a map on $\mathbb{R}^2$.

Proposition 36. *Complex inversion, translation and scaling are all orientation-preserving and conformal on $\mathbb{C}$.*

Proof. They are conformal and orientation preserving since they are holomorphic and have non-zero derivative away from ∞. □

Proposition 37. *Complex conjugation is conformal and orientation reversing on $\mathbb{C}$.*

Proof. As noted above, the Jacobian in question is

$$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

$\square$

We are interested in the notion of conformality and orientation-preservation/reversal on $\mathbb{P}^1(\mathbb{C})$. $\mathbb{P}^1(\mathbb{C})$ has two 'affine pieces', defined by

$$A_1 := \{[X, 1] : X \in \mathbb{C}\}, \quad A_2 := \{[1, Y] : Y \in \mathbb{C}\},$$

whose union covers all of $\mathbb{P}^1(\mathbb{C})$. The notion of conformality and orientation-preservation/reversal was defined on each such piece above. Therefore, to define it on $\mathbb{P}^1(\mathbb{C})$, we call on the local definition on any affine piece. It suffices to verify that the change of coordinates taking you from one piece to the other is holomorphic and hence conformal and orientation preserving. This change of coordinates is $z \mapsto 1/z$ (in both directions), which we have verified is holomorphic where it is defined, which is exactly on the overalap of the two pieces.

(It is satisfying to observe that more generally, we can define an affine piece formed of $\mathbb{P}^1(\mathbb{C})$ minus any point. The same will be true about all such change of coordinates maps.)

Theorem 38. *Circle inversions are orientation-reversing and conformal on $\mathbb{P}^1(\mathbb{C})$.*

Proof. We can show that the map is conformal on each of the affine pieces A_1 and A_2 discussed above. On each piece, which is a copy of $\mathbb{C}$, we have seen that circle inversion is of the form

$$z \mapsto \frac{a\overline{z} + b}{c\overline{z} + d}, \quad ad - bc \neq 0.$$

We have shown that such maps are compositions of complex conjugation with complex inversion, scaling and translations. The latter three types are conformal and orientation preserving. Complex conjugation is conformal and orientation reversing. By the previous propositions, the result is conformal and orientation reversing. $\square$

thm:mob **Theorem 39.** *The Möbius group of $\widehat{\mathbb{C}}$ is*

$$M(\widehat{\mathbb{C}}) = \left\{ z \mapsto \frac{az + b}{cz + d} : \begin{array}{l} a, b, c, d \in \mathbb{C} \\ ad - bc \neq 0 \end{array} \right\}.$$

Proof. We have seen that reflections are orientation reversing. We have also seen that the maps of the form above are all products of an even number of reflections. Conversely, we have also observed that a product of an even number of reflections is of the form above. $\square$

Theorem 40.
$$GM(\widehat{\mathbb{C}}) \cong M(\widehat{\mathbb{C}}) \rtimes \langle z \mapsto \overline{z} \rangle.$$

In particular, since $z \mapsto \overline{z}$ is of order two, $[GM(\widehat{\mathbb{C}}) : M(\widehat{\mathbb{C}}] = 2$.

Proof. Conjugation by $z \mapsto \overline{z}$ induces an automorphism of $M(\widehat{\mathbb{C}})$ inside $GM(\widehat{\mathbb{C}})$:

$$\left(z \mapsto \frac{az+b}{cz+d}\right) \mapsto \left(z \mapsto \frac{\overline{a}z+\overline{b}}{\overline{c}z+\overline{d}}\right).$$

Therefore $M(\widehat{\mathbb{C}})$ is normal in $GM(\widehat{\mathbb{C}})$. Further, we have seen that $GM(\widehat{\mathbb{C}})$ is generated by $M(\widehat{\mathbb{C}})$ and $z \mapsto \overline{z}$, and that $M(\widehat{\mathbb{C}})$ and $\langle z \mapsto \overline{z} \rangle$ are disjoint. $\square$

Stereographic projection, the chordal metric, and Möbius transformations can be defined in higher dimensions; see Beardon, Chapter 3.

3.5 The Action of $\mathrm{GL}_2(\mathbb{C})$ and $\mathrm{PGL}_2(\mathbb{C})$

We have now observed that $\mathrm{GL}_2(\mathbb{C})$ acts on $\widehat{\mathbb{C}}$ via Möbius transformation:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot z = \frac{az+b}{cz+d}$$

Proposition 41. *The kernel of this action is* $\{kI : k \in \mathbb{C}^*\}$.

Proof. A Möbius transformation
$$z \mapsto \frac{az+b}{cz+d}$$
acts trivially if and only if $cz^2 + (d-a)z - b = 0$ for all z, i.e. $c = b = 0$ and $d = a$. $\square$

We therefore define

$$\mathrm{PGL}_2(\mathbb{C}) = \mathrm{GL}_2(\mathbb{C})/\{kI : k \in \mathbb{C}^*\}.$$

This group acts faithfully on $\widehat{\mathbb{C}}$ by construction.

The map
$$\mathrm{GL}_2(\mathbb{C}) \to \mathrm{PGL}_2(\mathbb{C}),$$
is a homomorphism with kernel $\{kI : k \in \mathbb{C}^*\}$. Restricted to $\mathrm{SL}_2(\mathbb{C})$, it has kernel $\{I, -I\}$, and so we also define
$$\mathrm{PSL}_2(\mathbb{C}) := \mathrm{SL}_2(\mathbb{C})/\{\pm I\}.$$

Theorem 42. *There is a group isomorphism* $\mathrm{PGL}_2(\mathbb{C}) \cong M(\widehat{\mathbb{C}})$.

Proof. The map is $\mathrm{PGL}_2(\mathbb{C}) \to M(\widehat{\mathbb{C}})$ is injective by definition ($\mathrm{PGL}_2(\mathbb{C})$ is a quotient by the kernel of $\mathrm{GL}_2(\mathbb{C}) \to M(\widehat{\mathbb{C}})$). It is surjective by Theorem 39. $\square$

This tells us how to find the inverse of a Möbius function easily.

We have now observed that Möbius transformations are exactly the collection of linear automorphisms of $\mathbb{P}^1(\mathbb{C})$. In fact, all automorphisms of $\mathbb{P}^1(\mathbb{C})$ are of degree 1, so that one can define Möbius transformations as the automorphisms of $\mathbb{P}^1(\mathbb{C})$.

The functions

$$\frac{tr^{\,2}(A)}{\det(A)}, \quad \frac{||A||^2}{|\det(A)|}$$

on $\mathrm{GL}_2(\mathbb{C})$ are invariant under multiplication by $\mathbb{C}^*$, so that they are are well-defined on $\mathrm{PGL}_2(\mathbb{C}) \cong M(\mathbb{C})$. Therefore a Möbius transformation γ associated to $A \in \mathrm{PGL}_2(\mathbb{C})$ has a well-defined *square trace*,

$$tr^{\,2}(\gamma) = \frac{tr^{\,2}(A)}{\det(A)}$$

and *norm*,

$$||\gamma|| = \frac{||A||}{|det(A)|^{1/2}}.$$

This square trace is invariant under conjugation.

3.6 Cross-ratio

Theorem 43. *The* $\mathrm{PGL}_2(\mathbb{C})$ *action on* $\widehat{\mathbb{C}}$ *is simply (or 'sharply') triply transitive. In other words, if* $z_1, z_2, z_3, w_1, w_2, w_3 \in \widehat{\mathbb{C}}$*, then there exists a unique Möbius transformation* γ *such that* $\gamma(z_i) = w_i$ *for* $i = 1, 2, 3$*.*

Proof. Given $z_1, z_2, z_3 \in \widehat{\mathbb{C}}$, we will construct γ such that

$$\gamma(z_1) = 0, \gamma(z_2) = 1, \gamma(z_3) = \infty.$$

The needed transformation is

$$\gamma(z) = \frac{(z_2 - z_3)z - z_1(z_2 - z_3)}{(z_2 - z_1)z - z_3(z_2 - z_1)} = \frac{(z_2 - z_3)(z - z_1)}{(z_2 - z_1)(z - z_3)},$$

which can be verified directly. This gives existence. For uniqueness, we need only show that if $\gamma(0) = 0$, $\gamma(1) = 1$ and $\gamma(\infty) = \infty$, then $\gamma(z) = z$ for all z. Suppose

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

Then $\gamma(0) = 0$ implies that $b = 0$, $\gamma(1) = 1$ implies $a + b = c + d$, and $\gamma(\infty) = \infty$ implies $c = 0$. Therefore $a = d$ and we have

$$\gamma(z) = \frac{az}{a} = z.$$

□

Definition 44. *The cross-ratio of an ordered quadruple of distinct points* (z_1, z_2, z_3, z_4) *is*

$$(z_1, z_2; z_3, z_4) = \frac{(z_2 - z_3)(z_4 - z_1)}{(z_2 - z_1)(z_4 - z_3)}.$$

In other words, it is the value $\gamma(z_4)$ where γ is the unique Möbius transformation taking $z_1 \mapsto 0$, $z_2 \mapsto 1$, $z_3 \mapsto \infty$.

Note that there is no standard convention on the ordering of the points and the ordering of the ratios. Most authors disagree.

Of course, this involves a little finesse with ∞, for example,

$$(z_1, z_2; z_3, \infty) = \frac{(z_3 - z_2)}{(z_2 - z_1)}.$$

In essence, we cancel two ∞'s. If in doubt, return to the homogeneous representation:

$$\begin{aligned}&[(z_2/w_2 - z_3/w_3)(z_4/w_4 - z_1/w_1), (z_2/w_2 - z_1/w_1)(z_4/w_4 - z_3/w_3)] \\ &\quad = [(z_2w_3 - z_3w_2)(z_4w_1 - z_1w_4), (z_2w_1 - z_1w_2)(z_4w_3 - z_3w_4)].\end{aligned}$$

Setting $[z_4, w_4] = [1, 0]$, we obtain

$$[(z_2w_3 - z_3w_2)(w_1), (z_2w_1 - z_1w_2)(w_3)] = [z_2/w_2 - z_3/w_3, z_2/w_2 - z_1/w_1].$$

Next we will show that the cross-ratio is a 'projective invariant', i.e. invariant under Möbius transformation. In fact, we can characterize Möbius transformations by this property.

:crossinvariant **Proposition 45.** *A map* $\gamma : \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ *is a Möbius transformation if and only if it preserves cross-ratios, i.e.*

$$[z_1, z_2; z_3, z_4] = [\gamma(z_1), \gamma(z_2); \gamma(z_3), \gamma(z_4)].$$

Proof. Fix z_1, z_2, z_3. Write $w_i = \gamma(z_i)$. Then we have a diagram

$$\begin{tikzcd} \widehat{\mathbb{C}} \arrow[rr, "\gamma"'] \arrow[dr, "f_1"'] & & \widehat{\mathbb{C}} \arrow[dl, "f_2"] \\ & \widehat{\mathbb{C}} & \end{tikzcd}$$

where f_1 is the unique Möbius transformation that takes (z_1, z_2, z_3) to $(0, 1, \infty)$ and f_2 is the unique Möbius transformation that takes (w_1, w_2, w_3) to $(0, 1, \infty)$.

First, suppose γ preserves cross ratios. Then, in particular

$$f_1(z) = [z_1, z_2; z_3, z] = [w_1, w_2, w_3, \gamma(z)] = f_2(\gamma(z))$$

for all z, i.e., the diagram commutes. Hence $f_1 = f_2 \circ \gamma$ and so $\gamma = f_2^{-1} \circ f_1 \in M(\widehat{\mathbb{C}})$.

Conversely, if $\gamma \in M(\widehat{\mathbb{C}})$, then by the triply transitive property of Möbius transformations, a map taking (z_1, z_2, z_3) to $(0, 1, \infty)$ is unique. Therefore $f_1 = f_2 \circ \gamma$, and the diagram commutes. Since this holds for any choice of z_i, we have shown that γ preserves cross-ratios. ☐

Proposition 46. *The cross ratio satisfies*

$$(z_1, z_2; z_3, z_4) = (z_3, z_4; z_1, z_2) = (z_2, z_1; z_4, z_3) = (z_4, z_3; z_2, z_1).$$

These are all the permutations that fix no index.

Proposition 47. *If* $(z_1, z_2; z_3, z_4) = a$, *then*

1. $(z_1, z_3; z_2, z_4) = 1 - a$,
2. $(z_1, z_4; z_3, z_2) = 1/a$,
3. $(z_1, z_4; z_2, z_3) = 1/(1 - a)$,
4. $(z_1, z_2; z_4, z_3) = a/(a - 1)$,
5. $(z_1, z_3; z_4, z_2) = (a - 1)/a$.

These can be realized as the permutations that fix a given index, e.g. in the choices above, permutations fixing 1.

The subgroup of $M(\mathbb{C})$ given by

$$\left\{ z \mapsto z, \quad z \mapsto 1 - z, \quad z \mapsto \frac{1}{z}, \quad z \mapsto \frac{1}{1-z}, \quad z \mapsto \frac{z}{z-1}, \quad z \mapsto \frac{z-1}{z} \right\}$$

or, as matrices of $\mathrm{PGL}_2(\mathbb{C})$,

$$\left\{ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} -1 & 1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ -1 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 1 & -1 \end{pmatrix}, \begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix} \right\},$$

is called the *anharmonic group*, which acts on $\widehat{\mathbb{C}}$, permuting the set $\{0, 1, \infty\}$, and it is isomorphic to S_3. The two propositions above illustrate the exact sequence

$$1 \to V_4 \to S_4 \to S_3 \to 1.$$

3.7 The Möbius action on circles

Our goal is to show the following.

preservecircles **Theorem 48.** *Möbius transformations preserve circles.*

First, we record a corollary.

Corollary 49. *The $M(\widehat{\mathbb{C}})$ action on circles in $\widehat{\mathbb{C}}$ is transitive.*

Proof. A circle is determined by three points, so this follows from the fact that the $M(\widehat{\mathbb{C}})$ action on $\widehat{\mathbb{C}}$ is triply transitive. $\square$

Theorem 48 can be verified directly by breaking down Möbius transformations into their constituent parts and checking each directly, or calling on a result about circle inversions from plane geometry. We will use the Hermitian form definition of a circle (Definition 23). It will have the added benefit of giving a nice way to see the action of $\mathrm{PGL}_2(\mathbb{C})$ on the Hermitian form itself.

Proof. A circle is the zero locus of

$$(\overline{X} \quad \overline{Y}) \begin{pmatrix} a & b \\ \bar{b} & c \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix} = 0, \qquad a, c \in \mathbb{R}, b \in \mathbb{C}, ac - b\bar{b} < 0$$

that is,

$$\{\mathbf{v} \in \mathbb{P}^1(\mathbb{C}) : \mathbf{v}^* H \mathbf{v} = 0\}$$

where H is a Hermitian matrix (i.e. $H = H^* := \overline{H}^t$) of positive determinant.

Now, the action of $M \in \mathrm{PGL}_2(\mathbb{C})$ on this set takes it to

$$\{M\mathbf{v} \in \mathbb{P}^1(\mathbb{C}) : \mathbf{v}^* H \mathbf{v} = 0\} = \{\mathbf{v} \in \mathbb{P}^1(\mathbb{C}) : \mathbf{v}^* (M^{-1})^* H M^{-1} \mathbf{v} = 0\}$$

which is again the zero set of a new Hermitian matrix $(M^{-1})^* H M^{-1}$ of positive determinant. $\square$

Now let $\mathcal{H}$ denote the space of Hermitian matrices, which is a 4-real dimensional space. Circles are 1-dimensional subspaces of $\mathcal{H}$, i.e. elements of $\mathbb{P}^1(\mathcal{H})$. This observation motivates the consideration of the following action of $\mathrm{PGL}_2(\mathbb{C})$ on $\mathcal{H}$:

$$M \cdot H = (M^{-1})^* H M^{-1} |\det(M)|^2.$$

For example, let's consider the action of $z \mapsto iz + 1$ on the real line, $\widehat{\mathbb{R}}$, which is characterized by $X\overline{Y} = \overline{X}Y$, i.e. elements which are conjugate to themselves.

$$M = \begin{pmatrix} i & 1 \\ 0 & 1 \end{pmatrix}, \quad H = \begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix}.$$

Then

$$(M^{-1})^* H M^{-1} = \begin{pmatrix} 1 & 0 \\ -1 & -i \end{pmatrix} \begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 0 & i \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ -1 & 2 \end{pmatrix}.$$

This latter Hermitian form refers to $z + \overline{z} = 2$, i.e. $tr\ (z) = 1$, the vertical line through 1.

Some authors define instead

$$M \cdot H = \frac{MHM^*}{|\det(M)|^2}$$

which has the virtue of being simpler looking. Note that $M \mapsto (M^{-1})^*$ is an inversive automorphism of $\mathrm{PGL}_2(\mathbb{C})$; the actions differ by precomposition by this automorphism.

A corollary is the following.

prop:crossreal **Proposition 50.** *Four points lie on the same circle if and only if their cross ratio is real. Furthermore, z_4 and z_4^* are inverse points with respect to the circle generated by z_1, z_2, z_3 if and only if*

$$(z_1, z_2; z_3, z_4^*) = \overline{(z_1, z_2; z_3, z_4)}.$$

Proof. Map the circle generated by the first three points to $\widehat{\mathbb{R}}$ as in the proof of Proposition 45. Then, as in that proof, the fourth point is on the circle if and only if it maps to a real number; otherwise its inverse point with regards to $\widehat{\mathbb{R}}$ is its conjugate (details left to the reader). □

Exercise 51. *Do the details left out in the proof.*

Exercise 52. *Prove Theorem 48 by other means. For example, prove Proposition 50 first directly from the definition and then use it. Or use some other more or less conventional method.*

3.8 The inversive pairing on circles

The determinant is a well-defined function on $\mathcal{H}$ as a 4-dimensional vector space; in fact, it is a quadratic form

$$\det \begin{pmatrix} a & x+iy \\ x-iy & c \end{pmatrix} = ac - x^2 - y^2$$

of signature $(1,3)$, meaning it is conjugate to the (negative) Minkowski form

$$M(x, y, z, t) = t^2 - x^2 - y^2 - z^2.$$

This makes $\mathcal{H}$ into a vector space with a bilinear map $\langle \cdot, \cdot \rangle$ associated to the quadratic form

$$Q(a, c, x, y) = ac - x^2 - y^2.$$

The Gram matrix of this form is

$$G_Q := \begin{pmatrix} 0 & 1/2 & 0 & 0 \\ 1/2 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$$

defined by the property that

$$Q(a,c,x,y) = (a \quad c \quad x \quad y) G_Q \begin{pmatrix} a \\ c \\ x \\ y \end{pmatrix}.$$

Explicitly, the bilinear form is

$$\langle (a_1, c_1, x_1, y_1), (a_2, c_2, x_2, y_2) \rangle = \frac{1}{2} a_1 c_2 + \frac{1}{2} a_2 c_1 - x_1 x_2 - y_1 y_2.$$

The space $\mathcal{H}$ has an associated matrix group of linear transformations preserving this form, a type of orthogonal group:

$$O_Q(\mathbb{R}) := \{M \in M_4(\mathbb{R}) : M^t G_Q M = G_Q\}.$$

Note that the action of $\mathrm{PGL}_2(\mathbb{C})$ preserves the determinant. Therefore we have defined a homomorphism

$$\mathrm{PGL}_2(\mathbb{C}) \to O_Q(\mathbb{R}).$$

Definition 53. *Let C_1 and C_2 be two circles in $\widehat{\mathbb{C}}$, corresponding to vectors $\mathbf{x}_1$ and $\mathbf{x}_2 \in \mathcal{H}$, respectively, of length 1 under the bilinear form. The inversive product of these two circles is*

$$\Sigma(C_1, C_2) := \langle \mathbf{x}_1, \mathbf{x}_2 \rangle.$$

Since $\mathrm{PGL}_2(\mathbb{C})$ acts on $\mathcal{H}$ preserving determinant, we have the following fact.

Proposition 54. *Inversive product is preserved by Möbius transformations.*

3.9 Circles as elements of a Grassmanian

Note: I have not done this section in class, at the moment. Not sure if/when I will.

We can generalize projective space:

Definition 55. *For any field K, $\mathbb{P}^k(\mathbb{R})$ is defined as the set of k-dimensional subspaces of $\mathbb{R}^{k+1}$, given by*

$$\mathbb{P}^k(\mathbb{R}) := (\mathbb{R}^{k+1} \setminus \{\mathbf{0}\}) / \sim$$

where $\mathbf{v} \sim \mathbf{w}$ if and only if $\mathbf{v} = k\mathbf{w}$ for some $k \in K^*$.

We can generalize even further.

Definition 56. *The Grassmanian $\mathrm{Gr}_K(k, V)$ is the space of k-dimensional subspaces of a K-vector space V.*

Definition 57. *Two elements of* $\mathrm{Gr}_{\mathbb{R}}(2,\mathbb{C}^2)$ *are considered* $\mathbb{C}$*-equivalent if one is obtained from the other by multiplication by some* $\alpha \in \mathbb{C}^*$*. We will denote this by* $\sim_{\mathbb{C}}$*.*

Definition 58. *An element of* $\mathrm{Gr}_{\mathbb{R}}(2,\mathbb{C}^2)$ *is considered* $\mathbb{C}$*-trivial if its* $\mathbb{C}$*-span is a line. We will denote by* $\mathrm{Gr}_{\mathbb{R}}(2,\mathbb{C}^2)_0$ *the subset of non-*$\mathbb{C}$*-trivial elements.*

Proposition 59. *Let* C *be a circle of* $\mathbb{P}^1(\mathbb{C})$*. Then there exists a unique two-*$\mathbb{R}$*-dimensional non-*$\mathbb{C}$*-trivial subspace* $S(C)$ *of* $\mathbb{C}^2$*, up to* $\mathbb{C}$*-equivalence, such that* C *consists of the* $\mathbb{C}$*-lines of* $\mathbb{C}^2$ *intersecting* $S(C)$*. This association defines a bijection*

$$\left\{\textit{circles of } \mathbb{P}^1(\mathbb{C})\right\} \to \mathrm{Gr}_{\mathbb{R}}(2,\mathbb{C}^2)_0/\sim_{\mathbb{C}}$$

Proof. The elements (complex lines) of $\mathbb{P}^1(\mathbb{C})$ corresponding to $\widehat{\mathbb{R}}$ are exactly

$$\{[X,Y] : X/Y \in \mathbb{R} \text{ or } Y = 0\}$$

Therefore $\widehat{\mathbb{R}}$, as a circle of $\mathbb{P}^1(\mathbb{C})$, consists of all complex lines intersecting the non-$\mathbb{C}$-trivial real plane

$$\mathrm{Span}_{\mathbb{R}}\left\{\begin{pmatrix}0\\1\end{pmatrix}, \begin{pmatrix}1\\0\end{pmatrix}\right\},$$

or any other $\mathbb{C}$-equivalent plane. Furthermore, $\mathrm{PGL}_2(\mathbb{C})$ acts transitively on circles. Therefore, we need only show that $\mathrm{PGL}_2(\mathbb{C})$ acts transitively on $\mathrm{Gr}_{\mathbb{R}}(2,\mathbb{C}^2)_0/\sim_{\mathbb{C}}$ with the same stabilizer. The stabilizer of $\widehat{\mathbb{R}}$ under both actions is clearly $\mathrm{PGL}_2(\mathbb{R})$. For transitivity, we need only observe that all pairs of independent vectors in $\mathbb{C}^2$ are in the same orbit of $\mathrm{PGL}_2(\mathbb{C})$. □

To be concrete, suppose the element of $\mathrm{Gr}_{\mathbb{R}}(2,\mathbb{C}^2)$ in question is generated by

$$\mathbf{v} = \begin{pmatrix}\alpha\\\beta\end{pmatrix}, \quad \mathbf{w} = \begin{pmatrix}\gamma\\\delta\end{pmatrix}.$$

Then the corresponding circle is

$$i(\beta\overline{\delta} - \overline{\beta}\delta)X\overline{X} + i(\alpha\overline{\delta} - \gamma\overline{\beta})\overline{X}Y - i(\overline{\alpha}\delta - \overline{\gamma}\beta)X\overline{Y} + i(\alpha\overline{\gamma} - \overline{\alpha}\gamma)Y\overline{Y} = 0,$$

considered on $\mathbb{C}^2$. This can be verified by direct substitution. To go the other way, simply choose two distinct points of the circle C, and write them $\alpha/\beta,\quad \gamma/\delta$, in such a way that $(\alpha+\gamma)/(\beta+\delta)$ is also on the circle. Then the associated space is

$$\mathrm{Span}_{\mathbb{R}}\left\{\begin{pmatrix}\alpha\\\beta\end{pmatrix}, \begin{pmatrix}\gamma\\\delta\end{pmatrix}\right\}.$$

3.10 Topological action

Definition 60. *Suppose X is a topological space and $f : X \to Y$ is a surjection to a topological space Y. Then, f is a* quotient map, *and Y has the associated* quotient topology, *if the open sets of Y are exactly those sets $U \subset Y$ such that $f^{-1}(U)$ is open.*

This is the finest topology under which f is continuous.

prop:quotopen **Proposition 61.** *Let G be a group acting on a topological space X. Then the quotient map $f : X \to X/G$ by the action (i.e. identifying elements the share an orbit) is an open map.*

Proof. Let U be open in X. Then the preimage $f^{-1}(f(U))$ is exactly the orbit of U under G, which is a union of open sets gU, $g \in G$, hence open. Therefore, by the definition of quotient topology, $f(U)$ is open. □

Now we need a fact about when products and quotients commute.

prop:prodquot **Proposition 62.** *If $f : X \to A$ and $g : Y \to B$ are open quotient maps, then $f \times g : X \times Y \to A \times B$ is a quotient map.*

Proof. A product of continuous maps is continuous, so the preimages of open sets under $f \times g$ are open; the challenge is to show that having open preimage implies that a subset of $A \times B$ is open.

Suppose $W \subset A \times B$ has open preimage under $f \times g$. To show that W is open, it suffices to show it contains an open neighbourhood around each point $x \in W$. Choose $y \in (f \times g)^{-1}(x)$. Since $(f \times g)^{-1}(W)$ is open, and contains y, there is a basic open neighbourhood of y contained in $(f \times g)^{-1}(W)$, i.e. one the form $U \times V \subset X \times Y$, where U and V are open. Hence $f(U)$ and $g(V)$ are open (by f and g being open maps). Then the neighbourhood $f(U) \times g(V) \subset A \times B$ is an open neighbourhood of x contained in W. □

m:quotienttopgp **Theorem 63.** *Let H be a normal subgroup of a topological group G. Then G/H is a topological group under the quotient topology.*

Proof. We will essentially use a small lemma, as follow.

Lemma 64. *Given a commutative diagram of maps on topological spaces:*

$$\begin{tikzcd}
A \arrow[r, "f"'] \arrow[d, "g"'] & B \arrow[d, "h"'] \\
C \arrow[r, "j"'] & D
\end{tikzcd}$$

such that f, g and h are continuous, and furthermore, g is a quotient map, then j is continuous.

Proof. Let $U \subset D$ be open. Then the preimage $f^{-1} \circ h^{-1}(U)$ is open by continuity. But that implies $h^{-1}(U)$ is open by the fact that g is a quotient map. □

Returning to the proof, in the case of the inverse map, for example, we have a commutative diagram:

$$\begin{tikzcd} G \arrow[r] \arrow[d] & G \arrow[d] \\ G/H \arrow[r] & G/H \end{tikzcd}$$

The top and bottom arrows are $x \mapsto x^{-1}$. The left and right arrows are the quotient maps. Therefore by the lemma, the lower arrow is continuous.

In the case of multiplication, the challenge is in showing that the left map is a quotient map. We have a commutative diagram

$$\begin{tikzcd} G \times G \arrow[r] \arrow[d] & G \arrow[d] \\ G/H \times G/H \arrow[r] & G/H \end{tikzcd}$$

The top map is continuous by assumption. The right map is continuous by definition. The left-hand map is defined as a product of two quotient maps, each of which is open by Proposition 61. Therefore it is itself a quotient map, by Proposition 62. Therefore the lemma applies. □

quotientsection **Proposition 65.** *Let $f : X \to Y$ be a continuous surjective map of topological spaces, with a continuous section, i.e. a continuous map $g : Y \to X$ such that $f \circ g = id_Y$. Then f is a quotient map.*

Proof. Since f is continuous, it suffices to show that, for all $U \subset Y$, if $f^{-1}(U)$ is open, then U is open. Suppose $f^{-1}(U)$ is open. By continuity, $g^{-1} \circ f^{-1}(U) = U$ is open. □

Proposition 66. *The chordal metric topology on $\mathbb{P}^1(\mathbb{C})$ is the same as the quotient topology on $\mathbb{P}^1(\mathbb{C})$, considered as the quotient of $\mathbb{C}^2 \setminus \{(0,0)\}$ with the usual topology on $\mathbb{C}^2$, with quotient by the action of $\mathbb{C}^*$. Furthermore, the maps from the two standard affine pieces of $\mathbb{P}^1(\mathbb{C})$ to $\mathbb{C}$ are homeomorphisms.*

Proof. The overall plan is to consider a cover of $\mathbb{P}^1(\mathbb{C})$ by the sets obtained by deleting 0 and ∞ respectively. Since, under the quotient topology, or the chordal metric topology, these are open covers, it suffices to verify that the topologies agree on each piece.

Precisely, for each topology, we will show that the two pieces are both homeomorphic to $\mathbb{C}$ with the usual topology.

Removing the point $(0, 0, 1)$ from S^2, we obtain an open set stereographically projecting to $\mathbb{C}$ (corresponding to deleting ∞ from $\mathbb{P}^1(\mathbb{C})$). The chordal metric on S^2 induces a metric on $\mathbb{C}$ in this way. We will show this is the usual topology. A base of open sets for the chordal metric on $\mathbb{C}$ consists of sets of the form

$$U(a,r) := \left\{ z \in \mathbb{C} : \frac{|z-a|}{\sqrt{|z|^2+1}} < r \right\}$$

where $a \in \mathbb{C}$, $r \in \mathbb{R}$. A base of open sets for the usual topology on $\mathbb{C}$ is

$$B(a,r) := \{z \in \mathbb{C} : |z-a| < r\}$$

where $a \in \mathbb{C}$, $r \in \mathbb{R}$, Note that for any r,

$$B(a,r) \subset U(a,r)$$

And for any r,

$$U(a,r') \subset B(a,r)$$

where $r' = r/\sup_{z \in B(a,r)}\{\sqrt{|z|^2+1}\}$. Note that r' is well-defined and finite since $B(a,r)$ is bounded in the usual metric.

Therefore, given a point x contained in an open set U of one topology, there is a base open set V of the other topology so that $x \in V \subset U$. Therefore the chordal metric agrees with the usual topology on $\mathbb{C}$.

Now, if we remove instead the point $(0,0,-1)$, we can repeat the above computation. Looking at the homogeneous form of the metric, we see that the computation will run exactly the same (we dehomogenize at a different point, but obtain the same formula).

It remains, therefore, to check that the quotient topology on $\mathbb{P}^1(\mathbb{C})$ agrees with the usual topology on any affine piece. Consider the map

$$f : \mathbb{C}^2 \setminus \{(X,0) : X \in \mathbb{C}\} \to \mathbb{C}, \quad (X,Y) \mapsto X/Y,$$

as a map between two spaces with the usual $\mathbb{C}$-vector-space topology. The map is continuous and surjective from the formula. The section

$$g : \mathbb{C} \to \mathbb{C}^2 \setminus \{(X,0) : X \in \mathbb{C}\}, \quad z \mapsto (z,1)$$

is continuous. Therefore f is a quotient map, using Proposition 65.

Therefore the topology on $\mathbb{C}$ agrees with the quotient topology on the affine piece $\{[X,1] : X \in \mathbb{C}\}$. This works for the other affine piece in exactly the same way, and so the theorem is proven. $\square$

Exercise 67. *More generally, check that any affine piece is homeomorphic to* $\mathbb{C}$.

Definition 68. *Let G be a topological group and X be a topological space. An action of G on X is* continuous *if the action map $G \times X \to X$ is continuous.*

Theorem 69. *The group* $\mathrm{PGL}_2(\mathbb{C})$ *is a topological group under the quotient topology from* $\mathrm{GL}_2(\mathbb{C})$*, and its action on* $\widehat{\mathbb{C}}$ *is continuous.*

Proof. The group $\mathrm{PGL}_2(\mathbb{C})$, as a quotient of $\mathrm{GL}_2(\mathbb{C})$, has a quotient topology. By Theorem 63, the result is a topological group.

Consider also the topological space $\mathbb{C}^2 \setminus \{(0,0)\}$ with the usual topology restricted from $\mathbb{C}^2$.

We have a group action

$$\mathrm{GL}_2(\mathbb{C}) \times (\mathbb{C}^2 \setminus \{(0,0)\}) \to \mathbb{C}^2 \setminus \{(0,0)\}.$$

This action is continuous, since locally it looks like $\mathbb{C}^4 \times \mathbb{C}^2 \to \mathbb{C}^2$ given by

$$\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}, \begin{pmatrix} x \\ y \end{pmatrix} \right) \mapsto \begin{pmatrix} ax + by \\ cx + dy \end{pmatrix}.$$

Now we can take a quotient by the action of $\mathbb{C}^*$ on $\mathrm{GL}_2(\mathbb{C})$ and on $\mathbb{C}^2 \setminus \{(0,0)\}$. These quotients are open maps, by the proposition. We obtain two commuting group actions:

$$\begin{tikzcd}
\mathrm{GL}_2(\mathbb{C}) \times (\mathbb{C}^2 \setminus \{(0,0)\}) \arrow[r] \arrow[d] & \mathbb{C}^2 \setminus \{(0,0)\} \arrow[d] \\
\mathrm{PGL}_2(\mathbb{C}) \times \mathbb{P}^1(\mathbb{C}) \arrow[r] & \mathbb{P}^1(\mathbb{C})
\end{tikzcd}$$

We wish to show the lower arrow is continuous; the other three are known to be continuous.

If we have an open set of $\mathbb{P}^1(\mathbb{C})$ on the lower right, then by continuity, its preimage in the upper left is open. Hence its preimage in $\mathrm{PGL}_2(\mathbb{C}) \times \mathbb{P}^1(\mathbb{C})$ is open, since the left arrow is a quotient map by Proposition 62. □

3.11 Conjugacy classes: trace and fixed points

The goal of this section is to classify Möbius transformations up to conjugacy, and describe what the various classes do geometrically.

Proposition 70. *Suppose a group* G *acts on a set* X*. Then, if* $x, y \in X$ *are in the same orbit, then their stabilizers* G_x *and* G_y *are conjugate.*

Proof. Suppose $y = g \cdot x$. If $h \in G$ stabilizes x, then ghg^{-1} stabilizes y. Therefore $G_x \subset gG_yg^{-1}$. Similarly, since $x = g^{-1} \cdot y$, $G_y \subset g^{-1}G_xg$. Therefore, $G_x = gG_yg^{-1}$. □

Suppose a group G acts on a set X. Let $h \in G$. Then

$$hG_x = \{g \in G : g \cdot x = h \cdot x\}$$

More generally the left cosets of G_x are in bijection with the orbit of x.

Write F_g for the fixed points of g in X.

op:permutefixed **Proposition 71.** *Suppose a group G acts on a set X. Let $g, h \in G$. If $gh = hg$, then*

$$g \cdot F_h = F_h, \quad h \cdot F_g = F_g.$$

Note that this does not mean that g fixes the fixed points of h, but that it permutes them.

Actually, the statement above follows from a more general statement:

Proposition 72. *Suppose a group G acts on a set X. Let $g, h \in G$. Then*

$$F_{ghg^{-1}} = g \cdot F_h$$

Proof. Suppose $x \in F_h$. We wish to show $g \cdot x \in F_{ghg^{-1}}$. We compute

$$ghg^{-1} \cdot (g \cdot x) = gh \cdot x = g \cdot (h \cdot x) = g \cdot x.$$

Therefore $g \cdot F_h \subset F_{ghg^{-1}}$. Switching the roles of various players, we similarly obtain $g^{-1} \cdot F_{ghg^{-1}} \subset F_h$, which implies $F_{ghg^{-1}} \subset g \cdot F_h$. This completes the proof. □

In particular, the number of fixed points is a conjugacy invariant. However, we can find a finer invariant in the trace.

Definition 73. *The* square trace *of a Möbius transformation $M : z \mapsto \frac{az+b}{cz+d}$ is a complex number defined by*

$$tr\,^2(M) = \frac{(a+d)^2}{ad - bc}.$$

The trace *$tr\,(M)$ is defined as the equivalence class of the two square roots of the above. In other words, the trace is only defined up to sign.*

Beardon uses the square trace to classify Möbius transformations, and we will follow his lead, but be aware that many or most authors use the trace proper.

Note that the square trace is invariant under scaling a, b, c, d simultaneously by $\mathbb{C}^*$. Another way to compute the trace is to first normalize so that $ad - bc = 1$, and define it to be $\pm(a + d)$.

Proposition 74. *The square trace (and trace) is conjugacy invariant.*

This follows from the fact that trace and determinant of a matrix is invariant under conjugation.

We will follow Beardon's notation in defining the following standard Möbius transformations:

$$m_k : z \mapsto kz, \quad k \in \mathbb{C} \setminus \{0, 1\},$$

and

$$m_1 : z \mapsto z + 1.$$

Then, for all $k \neq 0$, including $k = 1$,

$$tr^{\,2}(m_k) = k + \frac{1}{k} + 2.$$

Note that m_1 has exactly one fixed point at ∞. The other m_k have two fixed points, 0 and ∞.

Theorem 75. *Two Möbius transformations are conjugate if and only if their square traces are equal. Furthermore, for every conjugacy class, there is a unique $k \in \mathbb{C}^*$ such that the class includes m_k and $m_{1/k}$.*

Proof. First, we will determine exactly when m_k and m_j may be conjugate. Since $tr^{\,2}$ is a conjugacy invariant, if m_k is conjugate to m_j, then

$$k + \frac{1}{k} = j + \frac{1}{j}$$

which occurs if and only if $k = j$ or $k = 1/j$. But m_k is conjugate to $m_{1/k}$ by $z \mapsto 1/z$. So m_k is conjugate to m_j if and only if $k = j$ or $k = 1/j$.

Let $f : z \mapsto \frac{az+b}{cz+d}$ be a Möbius transformation. It has at most two fixed points, since the fixed points are the solutions to a quadratic equation. Suppose first that it has two fixed points, say w_1 and w_2. Let h be any Möbius transformation taking

$$w_1 \mapsto 0, \quad w_2 \mapsto \infty.$$

Then $g = hfh^{-1}$ has fixed points 0 and ∞. Let $k = g(1)$. Then g and m_k agree on three points and therefore $g = m_k$.

Suppose instead that f has only one fixed point, w_1. Let $w_2 \neq w_1$. Let h be the unique Möbius transformation taking

$$w_1 \mapsto \infty, \quad w_2 \mapsto 0, \quad f(w_2) \mapsto 1$$

Then $g = hfh^{-1}$ has unique fixed point ∞ and satisfies $g(0) = 1$. The only Möbius transformations fixing ∞ and taking $0 \mapsto 1$ are of the form $z \mapsto az + 1$ where $a \in \mathbb{C}^*$. If $a \neq 1$, then such a transformation has two fixed points (the other being $1/(1-a)$). So $a = 1$ and $g = m_1$. $\square$

Exercise 76. *Determine the conjugacy classes of $\mathrm{PGL}_2(\mathbb{C})$ by linear algebraic methods, and show that you obtain the same answer as above. What do the eigenvectors and eigenvalues correspond to?*

It is convenient to classify Möbius transformations according to the classes to which they belong, i.e. according to their square trace. Each class exhibits certain characteristic geometric properties.

Proposition 77. *The Möbius transformation m_k is*

1. *a translation if and only if $k = 1$ if and only if $tr^2(m_k) = 4$*
2. *a rotation if and only if $|k| = 1$, $k \neq 1$ if and only if $tr^2(m_k) \in [0, 4)$*
3. *a dilation or contraction if and only if $k \neq 1$, $k \in \mathbb{R}^{>0}$ if and only if $tr^2(m_k) \in (4, \infty)$*

Note that each dilation is conjugate to a contraction and vice versa.

Proof. It is only the trace conditions that are not entirely obvious. For the first, it suffices to observe that $x + 1/x = 2$ has a double root $x = 1$. For the second, note that the unit circle under the function $x + 1/x$ maps 2-to-1 to the interval $[-2, 2]$. For the third, observe that the function $x + 1/x$ on $(0, \infty)$ is concave up with minimum 2 achieved at $x = 1$, which implies $tr^2 = z \in \mathbb{R}$ has two real solutions if and only if $z \geq 4$. □

Definition 78. *A Möbius transformation γ is*

1. parabolic *if it is conjugate to a translation,*
2. elliptic *if it is conjugate to a rotation,*
3. hyperbolic *if it is conjugate to a dilation,*
4. loxodromic *if it is not conjugate to a translation or rotation.*
5. strictly loxodromic *if it not conjugate to a translation, rotation or dilation.*

Using the previous proposition, then, we obtain an easy way to determine the type of a Möbius transformation:

Proposition 79. *A non-identity Möbius transformation γ is*

1. parabolic *if and only if $tr^2(\gamma) = 4$*
2. elliptic *if and only if $tr^2(\gamma) \in [0, 4)$*
3. hyperbolic *if and only if $tr^2(\gamma) \in (4, \infty)$*
4. loxodromic *if and only if $tr^2(\gamma) \notin [0, 4]$*
5. strictly loxodromic *if and only if $tr^2(\gamma) \notin [0, \infty)$*

Here I shall insert a nice Venn diagram.

$$VENN$$

Now, I wish to take a moment to discuss fixed points.

Definition 80. *Let $f : \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ be holomorphic. Let $\alpha \in \widehat{\mathbb{C}}$ be a periodic point of period n, i.e. $f^n(\alpha) = \alpha$.*

The point α is an attracting periodic point *if there is some open neighbourhood U of α such that f^{kn} uniformly converges to the constant function $f(z) = \alpha$ as $k \to \infty$.*

The point α is a repelling periodic point *if there is some open neighbourhood U of α such that for any neighbourhood V of α, for sufficiently large k, $U \subset f^{kn}(V)$.*

Otherwise α is called an indifferent *or* neutral periodic point*.*

op:attractrepel **Proposition 81.** *Let $f : \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ be holomorphic. Let $\alpha \in \mathbb{C}$ be a periodic point of period n, i.e. $f^n(\alpha) = \alpha$. Then α is attracting if and only if $|f'(\alpha)| < 1$, and it is repelling if and only if $|f'(\alpha)| > 1$.*

The quantity $f'(\alpha)$ is called the *multiplier* in the theory of dynamical systems.

Exercise 82.

1. *Show that the multiplier is invariant under conjugation of the map.*
2. *Extend the notion of multiplier to the appropriate definition at ∞, so that the above proposition still holds.*

attractrepelmob **Proposition 83.** *A non-identity Möbius transformation is*

1. *parabolic if and only if it has one fixed point*
2. *elliptic if and only if it has 2 neutral fixed points*
3. *loxodromic if and only if it has one attracting and one repelling fixed point*

Exercise 84. *Prove Propositions 81 and 83. You will have to have solved the previous exercise also (extending the multiplier).*

Note that the type of g is the same as the type of any iterate g^n of g.

Parabolic transformations. The map $z \mapsto z + 1$ keeps invariant all lines parallel to $\widehat{\mathbb{R}}$. It permutes the lines parallel to $i\widehat{\mathbb{R}}$. Each parabolic transformation with fixed point α has the same geometry: a family of circles tangent at α which are individually fixed, and an orthogonal family also tangent at α which are permuted. A family of circles tangent at one point is called a family of *horocycles*.

Hyperbolic transformations. The map $z \mapsto kz$, $k \in \mathbb{R}$, $k > 0$, $k \neq 1$, also keeps fixed a family of circles, this time the lines through 0 (i.e. circles through 0 and ∞). Orthogonal to these are the concentric family centred on 0; these are permuted. Again, any hyperbolic transformation will do something similar: it will fix circles passing through the two fixed points, and permute an orthogonal family (defined as those circles for which the two fixed points are inverse points).

Elliptic transformations. The map $z \mapsto e^{i\theta} z$ is described by the same two orthogonal families, except the first is permuted and the second is fixed. For any elliptic transformation, a similar geometry holds. The half turn $z \mapsto -z$ is special: it fixes the circles passing through 0 and ∞.

Loxodromic transformations. A loxodromic transformation with fixed points 0 and ∞ is of the form $z \mapsto ke^{i\theta}z$ where $k \in \mathbb{R}$, $k \neq 1$, $k > 0$. There are two invariant families – the same ones as for elliptic and hyperbolic transformations – but the circles of neither are fixed in general. In the hyperbolic case ($e^{i\theta} = 1$) the lines through the origin are fixed. These are also fixed in the *demi-hyperbolic* case ($e^{i\theta} = -1$). Loxodromic transformations also have other invariant curves: certain spirals around the fixed points.

Exercise 85. *In the geometric descriptions above, I've made all manner of assertions without proof. Prove the interesting ones.*

Exercise 86. *Classify commuting Möbius transformations.*

4 Hyperbolic geometry

In this section, I am in large part following a nice set of lecture notes due to John Parker, available at `maths.dur.ac.uk/~dma0jrp/img/HSjyvaskyla.pdf`.

4.1 Quaternion representation and Poincaré extension

The *Hamilton quaternions* over a field K, denoted $\mathbb{H}(K)$, is the ring whose additive structure is the four-dimensional K-vector space with basis $1, i, j, k$, and where the product is determined by the relations $i^2 = j^2 = -1$ and $k = ij = -ji$ and the assumption of associativity. This implies, in particular, $k^2 = -1$.

This ring is not commutative. It looks like

$$\mathbb{H}(K) = \{a + bi + cj + dk, \quad a, b, c, d \in K\}.$$

Exercise 87. *How is $\mathbb{H}(K)$ obtained from the quaternion group (Q_8 in Dummit and Foote)?*

We are interested in the case $K = \mathbb{R}$ for this section. We will just write $\mathbb{H} := \mathbb{H}(\mathbb{R})$. Then $\mathbb{C}$ is a subring, realized as those elements with $c = d = 0$. In fact,

$$\mathbb{H} = \{\alpha + \beta j, \quad \alpha, \beta \in \mathbb{C}\}.$$

This is because

$$a + bi + cj + dk = (a + bi) + (c + di)j.$$

Now, consider the complex plane as identified with the xy-plane in $\mathbb{R}^3$ with coordinates x, y, t. We will further identify (x, y, t) with $x + yi + tj \in \mathbb{H}$. The upper half space is then

$$H^3 := \{z + tj : z \in \mathbb{C}, t \in \mathbb{R}, t > 0\}.$$

It is open; its boundary is $\mathbb{C}$ itself ($t = 0$).

Fortunately for us, $\mathbb{H}$ is a division algebra.

Proposition 88. *Let $z \in \mathbb{C}$, $w \in \mathbb{C}$. Then,*

$$(z + wj)\left(\frac{\overline{z} - wj}{z\overline{z} + w\overline{w}}\right) = 1.$$

Proof. The proof is simply a computation, but it is a good exercise in non-commutative algebraic manipulations. First, it is useful to note that $jz = \overline{z}j$, since

$$jz = j(a + bi) = aj + bji = aj - bij = (a - bi)j = \overline{z}j.$$

From this, we compute

$$(z + wj)(\overline{z} - wj) = z\overline{z} + wj\overline{z} - zwj - wjwj = z\overline{z} + wzj - zwj - w\overline{w}j^2 = z\overline{z} + w\overline{w}$$

from which the result follows. $\square$

We can now extend Möbius transformations to act on H^3. Since $\mathbb{H}$ is not commutative, we choose to extend the Möbius transformation $\gamma : z \mapsto \frac{az+b}{cz+d}$ with $ad - bc = 1$ to

$$\gamma : z + tj \mapsto (a(z + tj) + b)(c(z + tj) + d)^{-1}.$$

We also extend $z \mapsto \overline{z}$ to

$$z + tj \mapsto \overline{z} + tj.$$

Since $\mathrm{PGL}_2(\mathbb{C}) \cong \mathrm{PSL}_2(\mathbb{C})$, i.e. any Möbius transformation can be scaled to give $ad - bc = 1$, we will only define the action on H^3 using such scaled Möbius transformations.

Proposition 89. *Möbius transformations take H^3 to itself.*

We could verify this directly for a general Möbius transformation, but it is perhaps worth it to see how the various types work out individually. To that end, let us do some computations:

Translation. The map $z \mapsto z + a$, $a \in \mathbb{C}$ acts by

$$z + tj \mapsto z + a + tj.$$

Scaling. The map $z \mapsto az$, $a \in \mathbb{C}^*$ must be realized as $z \mapsto \frac{sz+0}{0z+1/s}$ where $s^2 = a$. This acts by

$$z + tj \mapsto s(z + tj)s = szs + stjs = az + s\overline{s}tj = az + |a|tj.$$

Complex inversion. The map $z \mapsto 1/z$ should be realized as $z \mapsto \frac{i}{iz}$. This acts by

$$z + tj \mapsto i(i(z + tj))^{-1} = i(iz + tij)^{-1} = i\frac{-i\overline{z} - tij}{z\overline{z} + t^2} = \frac{\overline{z} + tj}{z\overline{z} + t^2}.$$

We will now record the general action of $z \mapsto \frac{az+b}{cz+d}$:

$$z + tj \mapsto \frac{(az + b)(\overline{cz} + \overline{d}) + a\overline{c}t^2 + tj}{|cz + d|^2 + |c|^2t^2}. \tag{1}$$

eqn:h3-mob

4.2 Hermitian forms

Definition 90. *Let V be a complex vector space. A* Hermitian form *on V is a map $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{C}$ which is complex linear in the first variable, and satisfies $\langle \mathbf{v}, \mathbf{w} \rangle = \overline{\langle \mathbf{w}, \mathbf{v} \rangle}$.*

These conditions entail that it is conjugate-linear in the second variable, i.e.

$$\langle \mathbf{v}, a\mathbf{w}_1 + b\mathbf{w}_2 \rangle = \overline{a}\langle \mathbf{v}, \mathbf{w}_1 \rangle + \overline{b}\langle \mathbf{v}, \mathbf{w}_2 \rangle.$$

The term *sesquilinear* refers to forms being linear in one variable and conjugate linear in the other; this is a weaker condition than being Hermitian.

Choosing a basis for V, any Hermitian form can be realized as

$$\langle \mathbf{v}, \mathbf{w} \rangle = \mathbf{w}^* H \mathbf{v}$$

where H is a Hermitian matrix, i.e. $H = H^*$. (Conversely, also, any Hermitian matrix gives a Hermitian form in this way.)

Exercise 91. *Prove the above.*

Proposition 92. *Let $\langle \cdot, \cdot \rangle$ be a Hermitian form. Then*

$$\mathbf{x} \mapsto \langle \mathbf{x}, \mathbf{x} \rangle$$

is a quadratic form on $\mathbb{C}$, considered as a real vector space, taking values in $\mathbb{R}$.

In particular, it is a function f satisfying the parallelogram law, $f(\mathbf{x}+\mathbf{y}) + f(\mathbf{x}-\mathbf{y}) = 2(f(\mathbf{x}) + f(\mathbf{y}))$, and $f(\alpha\mathbf{x}) = |\alpha|^2 f(\mathbf{x})$ for $\alpha \in \mathbb{C}$.

Definition 93. *The* discriminant *of a Hermitian form is the determinant of its Hermitian matrix.*

As with bilinear/quadratic forms, there are notions of definiteness and signature, etc.

Definition 94. *The Hermitian form associated to a Hermitian matrix H is* indefinite *if $\mathbf{x}^* H \mathbf{x}$ takes both positive and negative values. It is* positive definite *if it takes only positive values. It is* negative definite *if it takes only negative values.*

Two Hermitian matrices H_1 and H_2 are *congruent* if $H_2 = A^* H_1 A$ for some non-singular complex matrix A.

A form is called *degenerate* if $\det(H) = 0$. A Hermitian form induces a map from the vector space to its dual,

$$V^* = \{\mathbb{C}\text{-linear transformations } V \to \mathbb{C}\}.$$

The map is

$$V \to V^*, \quad \mathbf{x} \mapsto (\mathbf{y} \mapsto \langle \mathbf{y}, \mathbf{x} \rangle).$$

This map is an isomorphism if and only if there is no $\mathbf{x} \neq 0$ mapping to the zero map, if and only if $\det(H) \neq 0$ if and only if the form is non-degenerate.

Exercise 95. *Show that the eigenvalues of a Hermitian matrix are always real.*

The *signature* of a non-degenerate Hermitian form is (p, q) where p is the number of positive eigenvalues and q is the number of negative eigenvalues.

The theory of Hermitian forms is particularly simple, thanks to the following result, which is referred to as *Sylvester's law of inertia.*

Theorem 96. *The signature is a complete invariant for congruence classes of non-degenerate Hermitian forms.*

In particular, any Hermitian matrix is congruent to a diagonal one with entries ± 1. Two such matrices are congruent if and only if the number of positive (hence negative) entries is equal. Now the next result is clear.

Proposition 97. *A binary Hermitian form associated to matrix H is indefinite if and only if $\det(H) < 0$, and is definite if and only if $\det(H) > 0$.*

(Note the hypothesis that the form is binary, i.e. on a vector space of dimension two.)

Exercise 98. *Work out the analogous theory of bilinear forms on real vector spaces.*

Exercise 99. *Show that quadratic forms over $\mathbb{R}$ are classified up to congruence by their signature. Show that Hermitian forms over $\mathbb{C}$ are classified up to conjugacy by their signature (i.e. prove Sylvester's theorem).*

Given a non-degenerate Hermitian form on $\mathbb{C}^n$, there is an associated unitary group, i.e.

$$U(H) = \{A \in M_n(\mathbb{C}) : A^* H A = H\}.$$

The elements of $U(H)$ are distinguished by the property that they preserve the Hermitian form, i.e.

$$\langle A\mathbf{v}, A\mathbf{w} \rangle = \langle \mathbf{v}, \mathbf{w} \rangle.$$

Furthermore, every $A \in U(H)$ satisfies $|\det(A)|^2 = 1$. We also write

$$SU(H) = \{A \in U(H) : \det(A) = 1\}.$$

4.3 Hyperbolic functions

In a real vector space with the usual inner product, we have

$$\langle \mathbf{v}, \mathbf{w} \rangle = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle \langle \mathbf{w}, \mathbf{w} \rangle} \cos \theta$$

where θ is the angle between the vectors $\mathbf{v}$ and $\mathbf{w}$.

If we instead use a hyperbolic bilinear product such as

$$\langle (x_1, y_1), (x_2, y_2) \rangle = x_1 x_2 - y_1 y_1$$

on $\mathbb{R}^2$, then we come upon the hyperbolic cosine, cosh:

$$\langle \mathbf{v}, \mathbf{w} \rangle = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle \langle \mathbf{w}, \mathbf{w} \rangle} \cosh \theta$$

where θ now represents the 'hyperbolic angle'. The hyperbolic angle θ of a point on the unit hyperbola $x^2 - y^2 = 1$ is twice the area of the arc under the hyperbola swept out between the x-axis and the point. (This is in analogy to the fact that the angle between the axis and a ray from origin to a point P on the unit circle is equal to twice the area swept out under the unit circle from the x-axis to P.) The point then has coordinates

$$(\cosh \theta, \sinh \theta).$$

This motivates the notion of a 'distance' between linear subspaces of a vector space, given in terms of the bilinear pairing on the space, by

$$\frac{\langle \mathbf{v}, \mathbf{w} \rangle \langle \mathbf{w}, \mathbf{v} \rangle}{\langle \mathbf{v}, \mathbf{v} \rangle \langle \mathbf{w}, \mathbf{w} \rangle}.$$

The formula above is invariant under scaling either $\mathbf{v}$ or $\mathbf{w}$, so it is a well-defined measure of distance on linear subspaces. It is always a real number.

4.4 The upper half plane model of the hyperbolic plane

Before we put *geometry* on the upper half space, we are going to do the case of the upper half plane, because it is simpler and more familiar.

We have seen that $\widehat{\mathbb{C}}$ can be viewed as $\mathbb{P}^1(\mathbb{C})$, the quotient of $\mathbb{C}^2$. We will begin by putting a Hermitian structure on $\mathbb{C}^2$.

Now consider the Hermitian form

$$H = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix},$$

on the vector space $\mathbb{C}^2$. This Hermitian form has determinant -1, so it is indefinite. Explicitly,

$$\begin{pmatrix} \overline{z_1} & \overline{z_2} \end{pmatrix} \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = i z_2 \overline{z_1} - i \overline{z_2} z_1 = -2\Im(z_1 \overline{z_2}).$$

and

$$\begin{pmatrix} \overline{w_1} & \overline{w_2} \end{pmatrix} \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = i z_1 \overline{w_2} - i z_2 \overline{w_1}.$$

We write

$$\begin{aligned} \mathbb{C}^2_- &= \{\mathbf{v} \in \mathbb{C}^2 : \langle \mathbf{v}, \mathbf{v} \rangle < 0\} \\ \mathbb{C}^2_0 &= \{\mathbf{v} \in \mathbb{C}^2 : \langle \mathbf{v}, \mathbf{v} \rangle = 0\} \\ \mathbb{C}^2_+ &= \{\mathbf{v} \in \mathbb{C}^2 : \langle \mathbf{v}, \mathbf{v} \rangle > 0\} \end{aligned}$$

Note that $\Im(x\overline{y})$ and $\Im(x/y)$ have the same sign. Therefore, under the usual map to $\widehat{\mathbb{C}}$,

$$(x, y) \mapsto x/y$$

the image of $\mathbb{C}^2_0$ is $\widehat{\mathbb{R}}$, and the images of $\mathbb{C}^2_-$ and $\mathbb{C}^2_+$ are the upper and lower half planes, respectively.

Now we can give a metric on $\mathbb{C}^2_-$ according to the motivation of the last section.

Definition 100. *We define the* Poincaré metric $\rho(z, w)$ *on* z, w *in the upper half plane,* H^2*, by writing*

$$\cosh^2\left(\frac{\rho(z,w)}{2}\right) = \frac{\langle \mathbf{z}, \mathbf{w}\rangle\langle \mathbf{w}, \mathbf{z}\rangle}{\langle \mathbf{z}, \mathbf{z}\rangle\langle \mathbf{w}, \mathbf{w}\rangle},$$

where $\mathbf{z}$ *and* $\mathbf{w}$ *are any lifts of* z *and* w *from* $\widehat{\mathbb{C}}$ *to* $\mathbb{C}^2 \setminus \{(0,0)\}$*. This is invariant under the choice of lift.*

Proposition 101. *The associated line element* ds *satisfies*

$$(ds)^2 = 4\frac{\langle \mathbf{z}, d\mathbf{z}\rangle\langle d\mathbf{z}, \mathbf{z}\rangle - \langle d\mathbf{z}, d\mathbf{z}\rangle\langle \mathbf{z}, \mathbf{z}\rangle}{\langle \mathbf{z}, \mathbf{z}\rangle^2}.$$

Proof. We'd also like to derive the line element. The hyperbolic cosine has a Taylor expansion:

$$\cosh x = 1 + \frac{1}{2}x^2 + O(x^3).$$

We also make the simplifying assumption that $\mathbf{z}$ and $\mathbf{z} + d\mathbf{z}$ self-pair to -1, i.e.

$$\langle \mathbf{z} + d\mathbf{z}, \mathbf{z} + d\mathbf{z}\rangle = \langle \mathbf{z}, \mathbf{z}\rangle = -1$$

from which we may derive that

$$\langle d\mathbf{z}, d\mathbf{z}\rangle = -\langle \mathbf{z}, d\mathbf{z}\rangle - \langle d\mathbf{z}, \mathbf{z}\rangle.$$

Now we wish to compute $ds = \rho(\mathbf{z}, \mathbf{z} + d\mathbf{z})$. We will implicitly differentiate the simplified equation

$$\cosh\left(\frac{\rho(z,w)}{2}\right)^2 = \langle \mathbf{z}, \mathbf{w}\rangle\langle \mathbf{w}, \mathbf{z}\rangle,$$

obtaining, ignoring terms of higher order,

$$\begin{aligned} 1 + \frac{1}{4}(ds)^2 &= \langle \mathbf{z}, \mathbf{z} + d\mathbf{z}\rangle\langle \mathbf{z} + d\mathbf{z}, \mathbf{z}\rangle \\ &= \langle \mathbf{z}, \mathbf{z}\rangle + \langle \mathbf{z}, d\mathbf{z}\rangle + \langle d\mathbf{z}, \mathbf{z}\rangle + \langle \mathbf{z}, d\mathbf{z}\rangle\langle d\mathbf{z}, \mathbf{z}\rangle. \end{aligned}$$

Simplifying,

$$\frac{1}{4}(ds)^2 = -\langle d\mathbf{z}, d\mathbf{z}\rangle + \langle \mathbf{z}, d\mathbf{z}\rangle\langle d\mathbf{z}, \mathbf{z}\rangle.$$

Now, we will re-homogenize the equation, obtaining the theorem. $\square$

Using our particular choice of H, we obtain from the formulas above,

$$\cosh^2\left(\frac{\rho(z,w)}{2}\right) = \frac{|z-\overline{w}|^2}{4\Im(z)\Im(w)}.$$

An equivalent and useful formula is

$$\rho(z,w) = \log\frac{|z-\overline{w}|+|z-w|}{|z-\overline{w}|-|z-w|}.$$

This follows from the fact that

$$\cosh(z) = \frac{e^z+e^{-z}}{2}.$$

As for the differential, using the homogeneous form, we may choose any lift, so take $\mathbf{z} = (z,1)$ and $d\mathbf{z} = (dz,0)$. Then $\langle d\mathbf{z}, d\mathbf{z}\rangle = 0$, and the rest of the formula simplifies to

$$ds^2 = \frac{dz d\overline{z}}{\Im(z)^2} \quad \text{or} \quad ds = \frac{|dz|}{\Im(z)}.$$

It can be useful to write this in terms of $z = x + yi$:

$$(ds)^2 = \frac{(dx)^2+(dy)^2}{y^2}.$$

Exercise 102. *Use instead the Hermitian form* $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$. *What is different in the story above? We get a different model than the upper half plane model. Explain.*

Exercise 103. *It is possible to define the metric on the upper half plane by use of cross-ratios. How does that work?*

4.5 Isometries of the upper half plane

By construction, every element of $U(H)$ preserves the Poincaré metric, i.e. is a hyperbolic isometry. Elements of $U(H)$ are Möbius transformations when transported to $\widehat{\mathbb{C}}$: the question is which ones?

Proposition 104. *We have* $SU(H) = SL_2(\mathbb{R})$.

Proof. Set

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SU(H).$$

Then $ad - bc = 1$. Since H is invertible (*non-degenerate*), $A \in U(H)$ is equivalent to

$$A^{-1} = H^{-1}A^*H.$$

This becomes

$$\begin{pmatrix} d & -b \\ -c & a \end{pmatrix} = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \begin{pmatrix} \overline{a} & \overline{c} \\ \overline{b} & \overline{d} \end{pmatrix} \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} = \begin{pmatrix} \overline{d} & -\overline{b} \\ -\overline{c} & \overline{a} \end{pmatrix}$$

which occurs if and only if $a, b, c, d \in \mathbb{R}$. $\square$

There is another isometry besides these, which is the map

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

This map clearly preserves the upper half plane (which $z \mapsto \overline{z}$ does not). Although this is not accomplished by a matrix, it is true that it preserves the metric (though not the Hermitian form), since

$$\langle(-\overline{z}, 1), (-\overline{w}, 1)\rangle = \langle(w, 1), (z, 1)\rangle.$$

hm:isometriesh3 **Theorem 105.** *The isometries of the Poincaré upper half plane are* $\mathrm{PSL}_2(\mathbb{R}) \rtimes \langle z \mapsto -\overline{z}\rangle$.

Proof. We have verified that the stated group consists of isometries. It is direct to verify that conjugation by $z \mapsto -\overline{z}$ induces an automorphism of $\mathrm{PSL}_2(\mathbb{R})$. We must now show that this group is the full group of isometries. The proof will follow Parker's notes, Section 3.2, p. 18.

Suppose we have an isometry, ϕ. Suppose that $\phi(i) = a + bi$. By composing ϕ with a Möbius transformation with real coefficients, namely $z \mapsto \frac{z-a}{b}$, we obtain another isometry fixing i. Therefore let us assume ϕ fixes i.

Now, suppose $\phi(2i) = c + di$. Choosing a hyperbolic Möbius transformation γ with fixed points i and $-i$, we see that γ preserves $\widehat{\mathbb{R}}$ and H^2. As any γ preserves the circles orthogonal to the imaginary axis, we may choose an appropriate one which moves $c + di$ onto the imaginary axis. We may also assume it does so above i. Composing with ϕ, we may replace ϕ so that

$$\phi(i) = i, \phi(2i) = ti, \quad t > 1.$$

Now, some simple computations using the metric:

$$\cosh^2\left(\frac{\rho(2i, i)}{2}\right) = \frac{|3i|^2}{4\Im(2i)\Im(i)} = \frac{9}{8},$$
$$\cosh^2\left(\frac{\rho(ti, i)}{2}\right) = \frac{|(t+1)i|^2}{4\Im(ti)\Im(i)} = \frac{(t+1)^2}{4t}.$$

As ϕ is an isometry, these must be equal, which implies $t = 2$ or $t = 1/2$. As $t > 1$, we conclude $t = 2$. Therefore, ϕ fixes i and $2i$. Next we show that ϕ must fix ti for all $t > 0$.

Suppose $\phi(ti) = x + yi$, $y > 0$. Again we may compute various distances:

$$\cosh^2\left(\frac{\rho(ti,i)}{2}\right) = \frac{(t+1)^2}{4t},$$
$$\cosh^2\left(\frac{\rho(ti,2i)}{2}\right) = \frac{(t+2)^2}{8t},$$
$$\cosh^2\left(\frac{\rho(x+yi,i)}{2}\right) = \frac{x^2+(y+1)^2}{4y},$$
$$\cosh^2\left(\frac{\rho(x+yi,2i)}{2}\right) = \frac{x^2+(y+2)^2}{8y}.$$

Under the assumption that ϕ is an isometry, we obtain two equations which imply $x = 0$ and $y = t$. So ϕ fixes all of it, $t > 0$.

Finally, suppose $\phi(z) = w$. We wish to show that $z = w$ or $z = -\overline{w}$. We compare $\rho(z, yi)$ to $\rho(w, yi)$:

$$\frac{|z+yi|^2}{4y\Im(z)} = \frac{|w+yi|^2}{4y\Im(w)}.$$

Comparing coefficients of the powers of y, we obtain

$$\Im(w) = \Im(z), \quad \Im(w)z\overline{z} = \Im(z)w\overline{w}.$$

This implies that $z = w$ or $z = -\overline{w}$. By continuity, then ϕ is one of the two maps $\phi(z) = z$ or $\phi(z) = -\overline{z}$. $\square$

4.6 Geometry of H^2

Definition 106. *A* geodesic *is a line whose length is equal to the distance between its endpoints, or more generally, any curve such that any segment of the curve has this property.*

These are the correct analogue to the notion of 'straight line' in Euclidean geometry.

Proposition 107. *The geodesics of H^2 are exactly the arcs of circles orthogonal to $\widehat{\mathbb{R}}$.*

Proof. We will begin by computing the length of a segment of the imaginary axis. Write $z = pi$, $w = qi$, where $0 < p < q$. First, we compute

$$\cosh^2\left(\frac{\rho(z,w)}{2}\right) = \frac{|pi-qi|^2}{4pq} = \frac{(p+q)^2}{4pq}.$$

We also have

$$\cosh^2\left(\frac{\log(q/p)}{2}\right) = \left(\frac{\sqrt{p/q}+\sqrt{q/p}}{2}\right)^2 = \frac{(p+q)^2}{4pq},$$

which implies

$$\rho(z,w) = \log(q/p).$$

(Note that $\cosh^2$ is injective on positive real values.) Parametrize any curve γ joining z to w by

$$\gamma(t) = x(t) + y(t)i, \quad 0 \le t \le 1$$

where $\gamma(0) = z$, $\gamma(1) = w$. Then we may compute the length:

$$\begin{aligned}
||\gamma|| &= \int_\gamma \frac{|dz|}{\Im(z)} \\
&= \int_0^1 \frac{|x'(t) + y'(t)i|}{y(t)} dt \\
&\ge \int_0^1 \frac{|y'(t)i|}{y(t)} dt \\
&= \log(y(1)) - \log(y(0)) \\
&= \log(q/p) \\
&= \rho(z,w)
\end{aligned}$$

Furthermore, the inequality is strict if $x'(t) \neq 0$ for any t. Therefore, the only geodesic line joining z to w is the corresponding segment of the imaginary axis.

Now, consider two general points of the upper half plane, α and β. Then there is a Möbius transformation taking

$$i \mapsto \alpha, \quad ti \mapsto \beta$$

for some $t > 0$ and preserving the real line. To see this, for example, first translate left or right so that α moves to the imaginary axis, then dilate so that α moves to i. Finally, apply a hyperbolic isometry fixing i and $-i$ so that β moves to the imaginary axis.

Since Möbius transformations are isometries and preserve circles and angles, our result for the imaginary axis implies that there is a unique geodesic between α and β that is an arc of a circle orthogonal to the real axis.

Note that there is a unique circle containing two points and orthogonal to $\mathbb{R}$ (I will take this a fact from Euclidean geometry). □

We've actually shown something else in this proof:

Proposition 108. *The geodesics are the paths of shortest length between two points. Furthermore, the Poincaré metric is actually a metric.*

That is, we have shown that

$$\rho(z,w) = \inf ||\gamma||$$

where γ ranges over all curves which join z to w. In particular, this easily implies the metric axioms, including the triangle inequality. One could have defined the Poincaré metric this way, by first stipulating the line element.

Proposition 109. *Let z and w be two points in H^2. Let z^* and w^* be the boundary points of the infinite geodesic passing through z and w (where z^* is the boundary point closer to z, i.e. z^*, z, w, w^* is the order of the points along the circle). Then*

$$\rho(z, w) = \log[z, z^*, w, w^*].$$

Proof. We first verify this fact for $z = pi$ and $w = qi$, $0 < p < q$. Then $z^* = 0$ and $w^* = \infty$. Using the formula for cross-ratio,

$$[pi, 0; qi, \infty] = q/p.$$

This verifies the statement.

Therefore, the theorem follows from the fact that there is always a Möbius transformation taking any geodesic to the imaginary axis (as in the last proof). That is, if z and w are general upper half plane points, choose ϕ taking

$$z^* \mapsto 0, \quad z \mapsto pi, \quad w \mapsto qi, \quad w^* \mapsto \infty,$$

for some $0 < p < q$, and we obtain

$$\rho(z, w) = \rho(pi, qi) = \log[pi, 0; qi, \infty] = \log[z, z^*; w, w^*].$$

□

This is part of a larger story about cross-ratios: in fact, we can take the cross-ratio as the basis for defining hyperbolic geometry. Check out 'What is... a cross-ratio?' by Labourie for more info at `http://www.ams.org/notices/200810/tx081001234p.pdf`.

Exercise 110. *We have essentially derived the upper half plane model from the hyperboloid $\{\mathbf{x} \in \mathbb{C}^2 : \langle \mathbf{x}, \mathbf{x} \rangle = -1\}$, which is itself a model of the hyperbolic plane. Show that the geodesics in this model are intersections of this hyperboloid with certain planes in $\mathbb{C}^2$.*

Let us collect some facts about the geodesics in H^2, which are also called the *hyperbolic lines.*

Proposition 111.

1. *Any two points of H^2 determine a unique line.*
2. *Any two distinct lines intersect in at most one point.*
3. *There is an isometry taking any line to any other.*
4. *Reflection in a line is an isometry.*

5. *Given any line L and any point w, there is a unique line M through w and orthogonal to L.*

Proof. 1. We have seen this in the previous several proofs.

2. Lines are circles orthogonal to the real line; two circles intersect in at most two points, but as they are both orthogonal to the real line, these intersection points are inverse points with respect to the real line, so only one lies in H^2.

3. We have seen this in the previous several proofs.

4. Reflection in the imaginary axis is an isometry, combined with previous item.

5. Suppose L is the imaginary axis. Then the needed line is $M = \{z \in H^2 : |z| = |w|\}$. Now apply an isometry for the general case.

□

Definition 112. *We call two lines L and M* parallel *if they share an endpoint on the boundary $\widehat{\mathbb{R}}$. We call two non-parallel lines* disjoint *or* intersecting *if they are so in H^2.*

It is possible for non-parallel lines to be disjoint. Axiomatically, hyperbolic geometry differs from Euclidean geometry in that the parallel postulate

Postulate 113. *Given a line L and point P not on L, there* **is a unique line** *in the plane containing P and L which passes through P and does not intersect L.*

is replaced with

Postulate 114. *Given a line L and point P not on L, there* **are at least two lines** *in the plane containing P and L which pass through P and do not intersect L.*

Exercise 115. *Show that the hyperbolic plane is complete as a metric space.*

Exercise 116. *There are excellent opportunities for exercises or larger explorations in Albert Marden's book 'Outer Circles', Section 1.6*

4.7 Quaternionic Hermitian forms

The general theory of a Hermitian form on a complex vector space can be extended to a Hermitian form (in the quaternionic sense) on a vector space over $\mathbb{H}$.

We have *conjugation* in $\mathbb{H}$ defined by

$$z = a + bi + cj + dk \mapsto \overline{z} := a - bi - cj - dk$$

We define the *modulus* of $z \in \mathbb{H}$ to be $|z| = \sqrt{z\overline{z}}$. The *real part* of z is a, while the imaginary part is $bi + cj + dk$.

Proposition 117.
1. *We have* $|z| = \sqrt{a^2 + b^2 + c^2 + d^2} \in \mathbb{R}$.
2. *Complex conjugation satisfies* $\overline{(zw)} = \overline{w} \cdot \overline{z}$. *This means we must be careful with notation!*
3. *We have* $|zw| = |z||w|$.
4. *We have* $z^{-1} = \overline{z}|z|^{-2}$.
5. *We have* $(zw)^{-1} = w^{-1}z^{-1}$.

Proof. These are computational and not surprising. You might do a few to check. □

Definition 118. *A* quaternionic right vector space *is an abelian group* V *(under addition) with right scalar multiplication* $V \times \mathbb{H} \to V$ *that is associative and distributive, and where* $\mathbf{v}1 = \mathbf{v}$ *and* $\mathbf{v}0 = \mathbf{0}$ *(where* $\mathbf{0}$ *denotes the additive identity).*

Note that associativity is the fact that

$$(\mathbf{v} \cdot r) \cdot s = \mathbf{v} \cdot rs.$$

This is where the left vs. right really comes in. In other words, scalar multiplication is a *right action* on V.

Definition 119. *A* quaternionic Hermitian form *on a quaternionic right vector space* V *is a map* $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{H}$ *which is linear in the first variable and satisfies* $\langle \mathbf{v}, \mathbf{w} \rangle = \overline{\langle \mathbf{w}, \mathbf{v} \rangle}$.

As before, this implies $\langle \mathbf{v}, \mathbf{v} \rangle \in \mathbb{R}$.

Also as before, Hermitian forms are realized by Hermitian matrices, i.e. H satisfying $H^* = H$, where conjugation is quaternionic conjugation (note: this differs from the extension of complex conjugation $z \mapsto \overline{z}$ to an isometry of H^3), by the following formula:

$$\langle \mathbf{v}, \mathbf{w} \rangle = \mathbf{w}^* H \mathbf{v}.$$

4.8 Hyperbolic 3-space (upper half space model)

We have seen above how to define the upper half space as a subset of the quaternions. We have also seen the construction of the hyperbolic metric on the upper half plane. Now we construct the hyperbolic metric on the upper half space in a similar fashion.

We will consider the quaternionic right vector space $\mathbb{H}^2$. We will let

$$H = \begin{pmatrix} 0 & j \\ -j & 0 \end{pmatrix}.$$

This is a Hermitian matrix with quaternion entries.

We will now define the upper half space by considering the intersection of the light cone with V:

$$\begin{aligned} V_- &= \{\mathbf{v} \in V : \langle \mathbf{v}, \mathbf{v} \rangle < 0\} \\ V_0 &= \{\mathbf{v} \in V : \langle \mathbf{v}, \mathbf{v} \rangle = 0\} \\ V_+ &= \{\mathbf{v} \in V : \langle \mathbf{v}, \mathbf{v} \rangle > 0\} \end{aligned}$$

Now we wish to define a quaternionic projective space, so as to obtain a map from $\mathbb{H}^2 \setminus \{(0,0)\}$ to $\widehat{\mathbb{H}}$. Let

$$\mathbb{P}^1(\mathbb{H}) = \left(\mathbb{H}^2 \setminus \{(0,0)\}\right) / \sim$$

where $(\alpha, \beta) \sim (\gamma, \delta)$ if $(\alpha, \beta) = (\gamma, \delta)\kappa$ for some $\kappa \in \mathbb{H}^*$. Equivalently, if $\beta\alpha^{-1}\gamma\delta^{-1} = 1$. (We have $\kappa = \gamma^{-1}\alpha = \delta^{-1}\beta$.)

Then the map

$$\mathbb{P}^1(\mathbb{H}) \to \widehat{\mathbb{H}}, \quad [\alpha, \beta] \to \alpha\beta^{-1}$$

is well-defined, and, indeed, it is a bijection.

We will consider the associated quotient

$$\mathbb{H}^2 \setminus \{(0,0)\} \to \mathbb{H}, \quad (\alpha, \beta) \to \alpha\beta^{-1}.$$

Let $R = \mathbb{R} \oplus \mathbb{R}i \oplus \mathbb{R}j$. For the purpose of defining the upper half space, we will be interested in $\widehat{R}$ inside $\widehat{\mathbb{H}}$.

Now we wish to imitate the definition of the metric, but using this form. In order to do the computation to define the metric, we need a formula for this pairing. If we use a standard lift ($z_2 + t_2 j = w_2 + t_2 j = 1$), this is a bit more manageable:

$$\left\langle \begin{pmatrix} z + tj \\ 1 \end{pmatrix}, \begin{pmatrix} w + sj \\ 1 \end{pmatrix} \right\rangle = \begin{pmatrix} w + sj & 1 \end{pmatrix} \begin{pmatrix} 0 & j \\ -j & 0 \end{pmatrix} \begin{pmatrix} \overline{z} - tj \\ 1 \end{pmatrix} = -s - t + (w - z)\, j$$

And in the case that $z_1 + t_1 j = w_1 + s_1 j = z + tj$, we obtain

$$\begin{pmatrix} z + tj & 1 \end{pmatrix} \begin{pmatrix} 0 & j \\ -j & 0 \end{pmatrix} \begin{pmatrix} \overline{z} - tj \\ 1 \end{pmatrix} = -2t.$$

Note how similar this is to the upper-half plane computation.

In particular, the image of the subset V_-, intersected with R, becomes the upper half space in R, while V_0 becomes the complex plane.

Now we use this pairing to define the Poincaré metric on the upper half space, exactly as described for the upper half plane. We get

$$\cosh^2\left(\frac{\rho(z + tj, w + sj)}{2}\right) = \frac{|z - w|^2 + (t + s)^2}{4ts}, \quad ds^2 = \frac{dz\overline{dz} + dt^2}{t^2}.$$

If we write $z = z + yi$, we can write

$$ds^2 = \frac{dx^2 + dy^2 + dt^2}{t^2}.$$

4.9 Isometries of the Upper Half Space

Proposition 120. *Let* $H = \begin{pmatrix} 0 & j \\ -j & 0 \end{pmatrix}$. *Then we have*

$$\mathrm{SL}_2(\mathbb{C}) \subset U(H).$$

Proof. Direct computation:

$$\begin{aligned} A^* H A &= \begin{pmatrix} \overline{a} & \overline{c} \\ \overline{b} & \overline{d} \end{pmatrix} \begin{pmatrix} 0 & j \\ -j & 0 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \\ &= \begin{pmatrix} -\overline{c}j & \overline{a}j \\ -\overline{d}j & \overline{b}j \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \\ &= \begin{pmatrix} -jc & ja \\ -jd & jb \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \\ &= \begin{pmatrix} 0 & j \\ -j & 0 \end{pmatrix} \\ &= H. \end{aligned}$$

□

Therefore members of $\mathrm{SL}_2(\mathbb{C})$ preserve the Hermitian form. Now, observe that

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} z + tj \\ 1 \end{pmatrix} = \begin{pmatrix} az + atj + b \\ cz + ctj + d \end{pmatrix}.$$

In particular, the action of $\mathrm{PSL}_2(\mathbb{C})$ on $\mathbb{H}^2$ descends to the definition of Möbius transformation on H^2 that we have already defined, i.e.

$$z + tj \mapsto (az + atj + b)(cz + ctj + d)^{-1}.$$

Therefore we have shown that Möbius transformations are isometries of H^2.

Before we can determine the full set of isometries, we need to know a little more geometry of the Möbius action on H^3.

Translation. Consider the map $z \mapsto z + a$. As we saw, this extends to

$$z + tj \mapsto z + a + tj.$$

This clearly has only one fixed point, ∞, on the boundary of H^3 and no fixed points inside H^3.

Parabolic. More generally, an elliptic Möbius transformation will always have one fixed point on the boundary of H^3 and none inside.

Scaling. The map $z \mapsto az$, $a \in \mathbb{C}^*$ must be realized as $z \mapsto \frac{sz+0}{0z+1/s}$ where $s^2 = a$. This acts by

$$z + tj \mapsto s(z + tj)s = szs + stjs = az + s\bar{s}tj = az + |a|tj.$$

If $a \in \mathbb{R}^{>0}$, then the scaling is a dilation. It acts by

$$z + tj \mapsto a(z + tj)$$

which is just a dilation from the origin. The fixed points are 0 and ∞ and none inside H^3. On the other hand, if it is a rotation, i.e. $|a| = 1$, then we have

$$z + tj \mapsto az + tj$$

which means that it rotates the space around the axis tj, $t > 0$.

Hyperbolic. Elliptic transformations have two fixed points on the boundary of H^3 and none inside.

Elliptic. These transformations have a geodesic consisting entirely of fixed points joining the two fixed points of the boundary.

Strictly loxodromic. Strictly loxodromic transformations have two fixed points on the boundary.

Proposition 121. *The isometries of the Poincaré upper half space are* $\mathrm{PSL}_2(\mathbb{C}) \rtimes \langle z \mapsto \bar{z} \rangle$, *i.e.* $GM(\widehat{\mathbb{C}})$.

Proof. We have verified that the stated group consists of isometries. It is direct to verify that conjugation by $z \mapsto \bar{z}$ induces an automorphism of $\mathrm{PSL}_2(\mathbb{C})$. We must now show that this group is the full group of isometries.

Suppose we have an isometry, ϕ. Suppose that $\phi(j) = w + sj$. By composing ϕ with a Möbius transformation, namely translation by w followed by a dilation around 0, we obtain another isometry fixing j. Therefore let us assume ϕ fixes j.

Now, suppose $\phi(2j) = w + sj$. Choosing an elliptic Möbius transformation γ with fixed points i and $-i$, we see that γ fixes j, which is on the geodesic joining i and $-i$. By choosing an appropriate such transformation, we can move $w + sj$ onto the $\Re(z) = 0$ plane. Therefore we may assume $\Re(w) = 0$.

Now, choosing a elliptic transformation γ with fixed points 1 and -1, we see that γ fixes j as before. Since the geodesic from 1 to -1 is orthogonal to the $\Re(z) = 0$ plane, this elliptic transformation fixes that plane. Now we can choose γ so that $w + tj$ is moved onto the $\Im(z) = 0$ plane. Therefore we may assume $w = 0$, so that $\phi(2j) = tj$, for some $t > 0$. In fact, we may assume $t > 1$.

Now, almost exactly as in the proof for H^2, some simple computations using the metric:

$$\cosh^2\left(\frac{\rho(2i,i)}{2}\right) = \frac{9}{8},$$
$$\cosh^2\left(\frac{\rho(ti,i)}{2}\right) = \frac{(t+1)^2}{4t}.$$

As ϕ is an isometry, these must be equal, which implies $t = 2$ or $t = 1/2$. As $t > 1$, we conclude $t = 2$. Therefore, ϕ fixes j and $2j$. Next we show that ϕ must fix tj for all $t > 0$. Suppose $\phi(tj) = w + sj$, $s > 0$. Again we may compute various distances:

$$\cosh^2\left(\frac{\rho(tj, j)}{2}\right) = \frac{(t+1)^2}{4t},$$
$$\cosh^2\left(\frac{\rho(tj, 2j)}{2}\right) = \frac{(t+2)^2}{8t},$$
$$\cosh^2\left(\frac{\rho(w+sj, j)}{2}\right) = \frac{|w|^2 + (s+1)^2}{4s},$$
$$\cosh^2\left(\frac{\rho(w+sj, 2j)}{2}\right) = \frac{|w|^2 + (s+2)^2}{8s}.$$

Under the assumption that ϕ is an isometry, we obtain two equations which imply $|w| = 0$ and $s = t$. So ϕ fixes all of tj, $t > 0$.

Now the proof departs again from that of the H^2 case. Suppose $\phi(z + tj) = w + sj$. We compare $\rho(z + tj, yj)$ to $\rho(w + sj, yj)$:

$$\frac{|z|^2 + (t+y)^2}{4yt} = \frac{|w|^2 + (s+y)^2}{4ys}.$$

Comparing coefficients of the powers of y, we obtain

$$s = t, \quad |z|^2 = |w|^2.$$

In other words, ϕ fixes the complex modulus of the complex part, and the height above the complex plane. Therefore, composing with a rotation about the imaginary axis, we may assume, in addition to the previous assumptions, that ϕ fixes $1 + j$.

Now, the hyperbolic metric restricted to $\Im(z) = 0$ gives a copy of the upper half plane, in the sense that the Poincaré metric on H^3 restricts to the Poincaré metric on H^2. Therefore ϕ, restricted to $\Im(z) = 0$, is an isometry fixing $j\mathbb{R}$ and $1 + j$, hence by the classification of isometries of H^2, ϕ fixes all of $\Im(z) = 0$.

Now, for points of fixed j-height, the hyperbolic distance $\rho(z + tj, w + tj)$ is a fixed function of the complex distance $|z - w|$. In the complex plane, knowing the distance from a point z to every point of $\mathbb{R}$ determines z up to conjugation. Therefore, knowing the hyperbolic distance of $z + tj$ to every point of $\mathbb{R} + tj$ determines $z + tj$ up to conjugation.

Therefore, ϕ must take each $z + tj$ to either $z + tj$ or $\overline{z} + tj$. By continuity, ϕ must be the identity map or else $z + tj \mapsto \overline{z} + tj$.

In either case, ϕ is already in $\mathrm{PSL}_2(\mathbb{C}) \rtimes \langle z \mapsto \overline{z} \rangle$. □

4.10 Geometry of H^3

Proposition 122. *The extended Möbius transformations on H^3 are conformal and orientation preserving. Complex conjugation is conformal and orientation reversing.*

Proof. It is easy to verify this by computation for conjugation, translation, scaling and complex inversion. For example, the Jacobian matrix for the extension of scaling by a is block diagonal, with the first 2×2 block representing the Jacobian for its action $z \mapsto az$ on $\widehat{\mathbb{C}}$ and the second being the 1×1 block with entry $|a|$. The rest is left as an exercise for the reader. $\square$

I will take it as evident that circles in the complex plane are in bijection with hemispheres in H^3 orthogonal to the complex plane. Precisely, the boundary of the hemisphere is the circle. This is actually a statement about the intersection of spheres and planes in Euclidean space $\mathbb{R}^3$: that there is one and only one sphere orthogonal to a plane and intersecting in a specified circle. This bijection extends to lines in the complex plane, which correspond to planes in H^3 orthogonal to the boundary. We will also call these circles and hemispheres.

Proposition 123. *Let C be a circle in $\widehat{\mathbb{C}}$, given by the Hermitian form $\mathbf{x}^* H \mathbf{x} = 0$. Then the hemisphere above C is the set*

$$\left\{ \mathbf{x} = \begin{pmatrix} z + tj \\ 1 \end{pmatrix} : z \in \mathbb{C}, t \in \mathbb{R}, \mathbf{x}^* H \mathbf{x} = 0 \right\}. \tag{2}$$ eqn:hermhem

Proof. If one takes $H = \begin{pmatrix} a & b \\ \overline{b} & c \end{pmatrix}$, then (2) is of the form

$$z\overline{z}a + \overline{b}z + \overline{z}b + c = -at^2 \tag{3}$$ eqn:hemisphere

In other words, the intersection of the zero locus with the plane at height t above $\mathbb{C}$ is again a circle.

Let us consider lines and circles in $\mathbb{C}$ separately. The equation (3) intersects $t = 0$ in a circle when $a \neq 0$. In that case, a circle of the form

$$az\overline{z} + b\overline{z} + \overline{b}z + c = 0$$

can be rewritten

$$|z + b/a|^2 = \frac{b\overline{b} - ca}{a^2},$$

and therefore has centre b/a and radius squared $(b\overline{b} - ca)/a^2$. Therefore (2) consists of circles of radius squared $(b\overline{b} - ac)/a^2 - t^2$ at each height t. In other words, (2) is a hemisphere of radius squared $(b\overline{b} - ac)/a^2$.

The case of lines corresponds to the condition $a = 0$. Then the intersection with each plane parallel to $\mathbb{C}$ at height t is the same, and the Hermitian locus in the statement is a plane above the line, perpendicular to $\mathbb{C}$. $\square$

prop:moveshemi **Proposition 124.** *Suppose γ is a general Möbius transformation on $\widehat{\mathbb{C}}$, taking circle C_1 to circle C_2 in $\widehat{\mathbb{C}}$. Suppose hemispheres H_1 and H_2 lie above C_1 and C_2 respectively. Then the extension of γ to H^3 takes H_1 to H_2.*

Proof. This is now exactly as in the proof that Möbius transformations preserve circles. That is, the set

$$\{\mathbf{v} \in \mathbb{P}^1(\mathbb{H}) : \mathbf{v}^* H \mathbf{v} = 0\}$$

is taken to

$$\{\mathbf{v} \in \mathbb{P}^1(\mathbb{H}) : \mathbf{v}^* (M^{-1})^* H M^{-1} \mathbf{v} = 0\}.$$

and the statement of the proposition is obtained by restricting our consideration to $\mathbf{v} = [\alpha, \beta]$ such that $\alpha\beta^{-1} \in H^3$ (recall that Möbius transformations preserve the set of such $\mathbf{v}$). □

In particular, $M(\widehat{\mathbb{C}})$ is transitive on hemispheres.

Proposition 125. *The geodesics of H^3 are segments of lines or arcs of circles orthogonal to $\widehat{\mathbb{C}}$.*

Proof. First, we show that any segment of the j-axis is a geodesic. This computation proceeds very much like for H^2, but I will record it here. Write $z = pj$, $w = qj$, where $0 < p < q$. First, we compute

$$\cosh^2\left(\frac{\rho(z,w)}{2}\right) = \frac{(p-q)^2}{4pq}.$$

We implies, as before, that

$$\rho(z,w) = \log(q/p).$$

Parametrize any curve γ joining z to w by

$$\gamma(t) = x(t) + y(t)i + s(t)j, \quad 0 \le t \le 1$$

where $\gamma(0) = z$, $\gamma(1) = w$. Then we may compute the length:

$$\begin{aligned} ||\gamma|| &= \int_\gamma \frac{|dz|}{\Im(z)} \\ &= \int_0^1 \frac{|x'(t) + y'(t)i + z'(t)j|}{z(t)} dt \\ &\ge \int_0^1 \frac{|y'(t)|}{y(t)} dt \\ &= \log(y(1)) - \log(y(0)) \\ &= \log(q/p) \\ &= \rho(z,w) \end{aligned}$$

Furthermore, the inequality is strict if $x'(t) + y'(t)i \neq 0$ for any t. Therefore, the only geodesic line joining z to w is the corresponding segment of the imaginary axis.

Now we consider two arbitrary points α and β in H^3. There is a Möbius transformation taking $j \mapsto \alpha$ and $tj \mapsto \beta$ for some $t > 1$ (this is accomplished exactly as in the proof of Theorem 105).

Since $M(\widehat{\mathbb{C}})$ is conformal and takes hemispheres to hemispheres, this Möbius transformation moves the $\Re(z) = 0$ and $\Im(z) = 0$ planes to two orthogonally intersecting hemispheres, and the j-line to their intersection. The points α and β lie on this intersection. As an intersection of spheres in $\mathbb{R}^3$, this is a circle, itself orthogonal to $\mathbb{C}$.

Therefore the geodesic joining α to β is unique and is an arc of a circle orthogonal to $\mathbb{C}$ containing both points.

Note that there is a unique circle containing two points and orthogonal to $\mathbb{C}$ (I will take this a fact from Euclidean geometry). □

As before, we can now verify that ρ is a metric and geodesics are the shortest paths between points.

4.11 The Klein models

Let $\langle \cdot, \cdot \rangle$ be the non-degenerate quadratic form on $\mathbb{R}^{n+1}$ of signature $(n, 1)$ given by

$$x_1^2 + x_2^2 + \cdots + x_n^2 - x_{n+1}^2.$$

This form is realized as an orthogonal matrix, i.e.

$$\langle \mathbf{v}, \mathbf{w} \rangle = \mathbf{v}^t H \mathbf{w},$$

where $H^t = H$. We can, as before, define

$$\begin{aligned} V_- &= \{\mathbf{v} \in V : \langle \mathbf{v}, \mathbf{v} \rangle < 0\}, \\ V_0 &= \{\mathbf{v} \in V : \langle \mathbf{v}, \mathbf{v} \rangle = 0\}, \\ V_+ &= \{\mathbf{v} \in V : \langle \mathbf{v}, \mathbf{v} \rangle > 0\}. \end{aligned}$$

We can also define a projection map $\mathbb{R}^{n+1} \setminus \{(0,0)\} \to \mathbb{P}^n(\mathbb{R})$ by taking the quotient by the action of $\mathbb{R}^*$, as usual. In particular, $\mathbb{R}^n$ lives naturally inside $\mathbb{P}^n(\mathbb{R})$ as

$$(x_1, \ldots, x_n) \mapsto [x_1, \ldots, x_n, 1].$$

Then the Klein model is the projectivization of V_-, as before. This time, we will identify this with the set of vectors

$$\left\{(x_1, \ldots, x_n, 1) \in \mathbb{R}^{n+1} : x_1^2 + x_2^2 + \cdots x_n^2 - 1 < 0\right\}$$

In other words, the intersection of the interior of the light cone with a hyperplane at height 1.

In the case $n = 2$, we obtain a model of the hyperbolic plane called the Klein disc model. It has (by the same definitions in terms of the pairing as before):

$$\cosh^2\left(\rho((x_1, x_2), (y_1, y_2))\right) = \frac{(x_1y_1 + x_2y_2 - 1)^2}{(1 - x_1^2 - x_2^2)(1 - y_1^2 - y_2^2)},$$

and

$$ds^2 = \frac{dx_1^2 + dx_2^2 - (x_1dx_2 - x_2dx_1)^2}{(1 - x_1^2 - x_2^2)^2}.$$

There are corresponding formulas for hyperbolic n-space.

The geodesics in this model are straight lines in the Euclidean sense. More generally, the intersection of subspaces of $\mathbb{R}^{n+1}$ of dimension 1, 2 and 3 with the Klein model give points, geodesic lines, and geodesic planes, etc.

Exercise 126. *Prove the above.*

Considering the case of the Klein model of hyperbolic 3-space, the geodesic planes can be specified in terms of a normal vector in $\mathbb{R}^4$, which will have $\langle \mathbf{v}, \mathbf{v} \rangle > 0$.

We now wish to consider the transformations preserving the form, i.e.

$$\begin{aligned} O(n,1) &= \{A \in M_{n+1}(\mathbb{R}) : A^tHA = H\}, \\ SO(n,1) &= \{A \in O(n,1) : \det(A) = 1\}, \\ O^+(n,1) &= \{A \in O(n,1) : A \text{ is orthochronous}\}. \end{aligned}$$

Here, *orthochronous* means the transformation does not swap the two halves of the light cone. Note that any transformation in $O(n,1)$ satisfies $\det(A)^2 = 1$, so $\det(A) = \pm 1$. Elements of $SO(n,1)$ are called *proper*. When n is odd, these are the orientation preserving transformations.

These transformations of $\mathbb{R}^n$ preserve the form, and so they take V_- to V_-, and act on the Klein model. We write P in front of any matrix group to indicate we are considering it up to scaling; i.e., in the case of $O(n,1)$, up to multiplication by $\{\pm 1\}$. Since scaling does not affect the action on the Klein model, it is natural to consider $PO(n,1)$ in place of $O(n,1)$.

If n is even, then

$$PO(n,1) \cong SO(n,1) \cong PSO(n,1),$$

since scaling by -1 will change the sign of the determinant.

Whether n is even or odd,

$$PO(n,1) \cong O^+(n,1),$$

since scaling by -1 will change whether you are orthochronous or not. But if n is odd, then $PSO(n,1) \cong SO^+(n,1)$ is strictly smaller than $PO(n,1)$ (you can't change the sign of the determinant by scaling the matrix).

Proposition 127. *The full group of isometries of the Klein disc ($n = 2$) is $SO(2,1)$.*

In fact, the map $z \mapsto -\overline{z}$ on the upper half plane model, which is orientation reversing, corresponds to

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$

Exercise 128. *Find the map between the upper half plane model and the Klein disc model, and verify this claim.*

Proposition 129. *The full group of isometries of the Klein ball ($n = 3$) is $O^+(3,1)$. The orientation-preserving ones are $SO^+(3,1)$.*

In particular, this implies there is an isomorphism

$$\mathrm{PSL}_2(\mathbb{C}) \to PSO(3,1) \cong SO^+(3,1).$$

Finally, I will mention the hyperboloid model, which is also formed from $\mathbb{R}^n$ but by choosing a lift of the projectivization of the light cone to the hyperboloid $\{\mathbf{v} \in \mathbb{R}^n : \langle \mathbf{v}, \mathbf{v} \rangle = -1\}$. In this case, we get the pleasingly simple

$$ds^2 = dx_1^2 + dx_2^2 + \cdots + dx_n^2 - dx_{n+1}^2.$$

Again, geodesic lines and planes are given by intersections of 2- and 3-dimensional subspaces of $\mathbb{R}^4$. The isometries are as above.

Exercise 130. *Any of the Exercises 1.1 of Maclachlan and Reid. (Careful which model you're using.)*

4.12 Hyperbolic geometry in the plane: area, triangles, etc.

Proposition 131. *Hyperbolic circles in H^2 coincide with Euclidean circles. In particular, the hyperbolic metric induces the usual topology on H^2 as a subset of the complex plane.*

Proof. The condition $\rho(z, w) < k$ on z becomes a condition of the form

$$|z - \overline{w}|^2 < c\Im(z)$$

which is the equation of a disc, albeit not centred at $\overline{w}$. Therefore the two topologies have the same basis of open sets. ☐

The area element associated to the hyperbolic metric on H^2 is computed from the line element, and it is:

$$dA = \frac{dx\, dy}{y^2}.$$

There is also a volume element for hyperbolic 3-space, as follows:

$$dV = \frac{dx\, dy\, dt}{t^2}.$$

By definition isometries preserve area and volume.

The data of a metric also defines *angles*. That is, the metric gives a pairing on the tangent space at a point z (a Riemannian metric). This pairing can also be derived from the line element:

$$\langle \mathbf{v}, \mathbf{w} \rangle = \frac{v_1 w_1 + v_2 w_2}{y^2}.$$

Then the angle between two geodesics is the angle between their tangent vectors, according to the pairing above.

Proposition 132. *Hyperbolic angles coincide with Euclidean angles in H^2.*

Proof. The Euclidean metric has pairing:

$$\langle \mathbf{v}, \mathbf{w} \rangle_E = v_1 w_1 + v_2 w_2.$$

In particular, at any point,

$$\frac{\langle \mathbf{v}, \mathbf{w} \rangle^2}{\langle \mathbf{v}, \mathbf{v} \rangle \langle \mathbf{w}, \mathbf{w} \rangle} = \frac{\langle \mathbf{v}, \mathbf{w} \rangle_E^2}{\langle \mathbf{v}, \mathbf{v} \rangle_E \langle \mathbf{w}, \mathbf{w} \rangle_E}$$

which implies that angles coincide. □

It is natural to define the angle between two parallel geodesics to be 0.

Exercise 133. *Decide what 'natural' means here: which facts hold with this definition that don't hold otherwise? Prove it.*

Hyperbolic polygons are simply-connected closed sets whose boundaries consist of a finite number of geodesic segments. The vertices may lie in H^2 or on the boundary (i.e. where two parallel geodesics meet at ∞).

Theorem 134 (Lambert, special case of Gauss-Bonnet)**.** *A hyperbolic triangle with angles θ_1, θ_2 and θ_3 has area $\pi - \theta_1 - \theta_2 - \theta_3$.*

Proof. This standard proof follows, for example, Katok, *Fuchsian Groups*, Theorem 1.4.2, and I refer the reader to it for figures since I can't LaTeX them here.

First we assume that the triangle in question is of the following form: two sides are vertical lines above real values in the interval $[-1, 1]$, and one side is an arc of the unit circle. Then, for some $\alpha, \beta \in [0, \pi/2]$, the left and right vertices of the triangle lie at $(\cos(\pi - \alpha), \sin(\pi - \alpha))$ and $(\cos\beta, \sin\beta)$, respectively.

Now, the interior angles at the two vertices in H^2 are exactly α and β (by a high school geometry proof).

Now we compute the area:

$$\int_A \frac{dx\, dy}{y^2} = \int_a^b dx \int_{\sqrt{1-x^2}}^{\infty} \frac{dy}{y^2} = -\int_{\pi-\alpha}^{\beta} d\theta = \pi - \alpha - \beta.$$

where we have substituted $x = \cos\theta$.

Any triangle which has at least one vertex on the boundary of H^2 can be moved, by a Möbius transformation of $\mathrm{PSL}_2(\mathbb{R})$, to the triangle just considered. For, we can first move the boundary vertex to ∞ by an element of $\mathrm{PSL}_2(\mathbb{R})$. Then the finite geodesic side will be an arc of a circle. By dilation and translation, we can move this to the unit circle.

Now, suppose we wish to consider a triangle whose vertices all lie in H^2. Then it can be obtained as the difference between two triangles with at least one vertex on the boundary each (picture in Katok; extend one geodesic side to the boundary). Then the area is the difference of the two areas; labelling all the angles (keeping in mind the angle is 0 at the boundary, and that a particular pair of angles will sum to π), we obtain the desired result. $\square$

As a corollary, the angle sum of any hyperbolic triangle is less than π.

More generally,

Theorem 135. *If any hyperbolic polygon has interior angles $\alpha_1, \ldots, \alpha_n$, then its hyperbolic area is $(n-2)\pi - \sum_{i=1}^{n} \alpha_i$.*

Exercise 136. *Prove this. Hint: non-convex polyhedra are harder to triangulate, but any triangulation will satisfy Euler's formula. (Solution in Beardon, Theorem 7.15.1.)*

I will now record some hyperbolic trigonometrical statements. I will not provide proof, but refer the reader to Katok, *Fuchsian groups*, Section 1.5.

Proposition 137. *Consider a hyperbolic triangle of angles α, β and γ and opposite sides a, b and c. Then we have the following:*

1. *Sine Rule:*
$$\frac{\sinh a}{\sin\alpha} = \frac{\sinh b}{\sin\beta} = \frac{\sinh c}{\sin\gamma}$$

2. *Cosine Rule I:*
$$\cosh c = \cosh a \cosh b - \sinh a \sinh b \cos \gamma.$$
3. *Cosine Rule II:*
$$\cosh c = \frac{\cos \alpha \cos \beta + \cos \gamma}{\sin \alpha \sin \beta}.$$

Corollary 138. *There is an isometry mapping between any two triangles of the same angles.*

Proof. By Cosine Rule II, the side lengths of the triangle are determined entirely by its angles. Therefore the sides match up in length.

We have seen already (proof of the classification of geodesics) that we can map any pair of points x, y separated by a given distance to any other pair x', y' separated by the same distance. Therefore we can map by isometry one side, say a joining x and y, of the first triangle Δ to a corresponding side, say a' joining x' and y', of the second triangle Δ'.

Finally, the third point of Δ, call it z, must end up, under this isometry, at one of at most two possible points. To see this, draw the hyperbolic circles centred x' and y'; they are also Euclidean circles and intersect at most twice. One of these points is the third vertex, z' of Δ'. These two points are related by reflection in the side a', which is an isometry. Therefore we can map the first triangle to the second. □

Exercise 139. *Show that a hyperbolic circle of radius r has circumference $2\pi \sinh r$. Show that the area of the hyperbolic disc of radius r is $4\pi \sinh^2(r/2)$.*

Exercise 140. *Show that H^2 and H^3 are complete but not compact.*

Exercise 141. *Define a notion of dihedral angle and investigate its properties in hyperbolic geometry.*

5 Subgroups of $\mathrm{PSL}_2(\mathbb{C})$

5.1 Fixed points and stabilizers for $\mathrm{PSL}_2(\mathbb{C})$

Write
$$[g, h] = ghg^{-1}h^{-1}.$$

For a Möbius transformation, trace is only determined up to sign, since one may multiply the transformation by ± 1. But $tr\ [g, h]$ can be uniquely determined by taking it to be $ghg^{-1}h^{-1}$ itself, not just the $\mathrm{PSL}_2(\mathbb{C})$ equivalence class; this is invariant under multiplying g or h by -1.

Note: This is a really strange thing to do! We are not considering $[g, h]$ as an element of $\mathrm{PSL}_2(\mathbb{C})$, but as an expression in such elements. This is related to the interesting fact that the exact sequence

$$1 \longrightarrow \{\pm 1\} \longrightarrow \mathrm{SL}_2(\mathbb{R}) \longrightarrow \mathrm{PSL}_2(\mathbb{R}) \longrightarrow 1$$

does not split and yet discrete subgroups of $\mathrm{PSL}_2(\mathbb{R})$ can always be lifted.

prop:traceshare **Proposition 142.** *Let $g, h \in \mathrm{PSL}_2(\mathbb{C})$ be non-identity transformations. Then g and h share a common fixed point as Möbius transformations if and only if $tr\ [g, h] = 2$.*

Proof. The following is an exact sequence of groups:

$$1 \longrightarrow \left\{ \begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix} : a \in \mathbb{C} \right\} \longrightarrow \left\{ \begin{pmatrix} s & a \\ 0 & s^{-1} \end{pmatrix} : s \in \mathbb{C}^*, a \in \mathbb{C} \right\} \longrightarrow \mathbb{C}^* \longrightarrow 1\ ,$$

where the third arrow is given by

$$\begin{pmatrix} s & a \\ 0 & s^{-1} \end{pmatrix} \mapsto s.$$

The kernel is characterized as the elements of trace exactly 2. Since $\mathbb{C}^*$ is abelian, every commutator of elements fixing ∞ is in this kernel. Since any two elements sharing a fixed point can be conjugated so the shared point is ∞, we have shown that sharing a fixed point implies $tr\ [g, h] = 2$.

Now, conversely, suppose $tr\ [g, h] = 2$.

Case I: Suppose g fixes exactly two points. Then, conjugating, g can be written diagonally (i.e. fixing 0 and ∞), and h written generally:

$$g = \begin{pmatrix} k & 0 \\ 0 & k^{-1} \end{pmatrix}, \quad h = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

Then $2(ad - bc) = 2 = tr\ [g, h] = 2ad - bc(k^2 + k^{-2})$, which implies that $k = 1$ or $bc = 0$. Hence g is the identity map (a contradiction) or h fixes 0 and/or ∞, sharing a fixed point with g.

Case II: Now suppose that g is parabolic. Then we can conjugate g to fix ∞ and be of the form

$$g = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix},$$

whence $tr\ [g, h] = 2 + c^2$. Therefore $c = 0$ and h fixes ∞ along with g. Hence g and h share exactly one fixed point. $\square$

If one chooses lifts $\overline{g}, \overline{h} \in \mathrm{SL}_2(\mathbb{C})$ and puts the matrices id, $\overline{g}$, $\overline{h}$, and $\overline{g}\overline{h}$ in a matrix $M_{g,h}$ as columns, then

$$\det M_{g,h} = 2 - tr\,[g,h].$$

Therefore we can characterize when g and h share a fixed point as a question about linear independence of the matrices id, g, h, gh.

We call the geodesic joining the two fixed points of a non-parabolic, non-identity map the *axis*.

eclassification **Proposition 143.** *The following are equivalent, for non-identity Möbius transformations g and h.*

1. *g and h commute.*
2. *one of the following holds:*
   (a) *$F_g = F_h$*
   (b) *each of g and h are involutions, each interchanging the fixed points of the other.*
3. *one of the following holds:*
   (a) *g and h are parabolic with the same fixed point,*
   (b) *g and h have the same axis,*
   (c) *g and h are involutions, with orthogonally intersecting axes.*

Proof. Suppose that g and h commute. By Proposition 71, g and h each act by a permutation on each others' fixed points.

Suppose that $|F_g| = |F_h| = 2$. Then, if g fixes F_h pointwise, we have $F_g = F_h$. Otherwise, g and h interchange the fixed points of the other, but do not share fixed points. In this case, g^2 has four fixed points, so $g^2 = id$. Similarly, $h^2 = id$. Now, if F_g and F_h are not both of size two, then without loss of generality, suppose g is parabolic (having one fixed point). Then h fixes F_g, so they share a fixed point and in fact $F_g = F_h$ of size 1.

This shows (1) implies (2).

If $F_g = F_h$ of size 1, then g and h are parabolic with the same fixed point.

If $F_g = F_h$ of size 2, then g and h have the same axis.

If g and h are involutions, each interchanging the fixed points of the other, then let us conjugate and consider g to have fixed points 0 and ∞, hence with axis equal to the j-axis. Then, as g is an involution, it must be the map $z \mapsto -z$. If h is elliptic and interchanges 0 and ∞, then it must be of the form $z \mapsto a^2/z$ for some $a \in \mathbb{C}^*$. Then its fixed points are $\pm a$, and the axis between them intersects the j-axis orthogonally.

This shows (2) implies (3).

Now, any two parabolics with the same fixed point commute, as do any transformations sharing an axis (to see this, conjugate so the fixed points are ∞ in the first case and 0 and ∞

in the second, and examine the matrix form). This leaves the third case for consideration. Suppose g and h are involutions with orthgonally intersecting axes. By conjugation, we may assume the maps are $z \mapsto -z$ and $z \mapsto a^2/z$, which commute. $\square$

p:ellipticfixed **Proposition 144.** *A group of Möbius transformations contains only elliptic elements if and only if all elements share a common fixed point in H^3. In particular, a finite group has a common fixed point in H^3.*

The proof of this is a little long, but geometrically interesting. It is in Beardon, Theorem 4.3.7.

5.2 Point stabilizers

We will now discuss the unitary group in $GL_2(\mathbb{C})$, which is simply the unitary group with respect to the Hermitian pairing whose Gram matrix is the identity:

$$\begin{aligned}
\mathrm{U}_2(\mathbb{C}) &= \{M \in \mathrm{GL}_2(\mathbb{C}) : M^*M = I\}, \\
\mathrm{SU}_2(\mathbb{C}) &= \{M \in \mathrm{SL}_2(\mathbb{C}) : M^*M = I\}, \\
\mathrm{PSU}_2(\mathbb{C}) &= \{M \in \mathrm{PSL}_2(\mathbb{C}) : M^*M = I\}.
\end{aligned}$$

These represent 'complex rotations' in analogy to orthogonal matrices representing rotations of real vector spaces. That is, they preserve distances and angles in $\mathbb{C}^2$ defined by the usual complex inner product $\langle \mathbf{v}, \mathbf{w} \rangle = v_1\overline{w_1} + v_2\overline{w_2}$. In other words, $SU_2(\mathbb{C})$ represents rotations of S^3 in $\mathbb{C}^2 \cong \mathbb{R}^4$. The projectivization $\mathbb{C}^2 \setminus \{(0,0)\} \to \mathbb{P}^1(\mathbb{C})$, applied to S^3 is the Hopf fibration,

$$1 \longrightarrow S^1 \longrightarrow S^3 \longrightarrow S^2 \longrightarrow 1$$

whose fibres are circles. The S^2 here lies in $\mathbb{R}^3$. This is one justification for the isomorphism

$$\mathrm{PSU}_2(\mathbb{C}) \cong SO(3),$$

where the latter is the special orthogonal group on $\mathbb{R}^3$, i.e. transformations preserving length and orientation: the rotations of S^2 in $\mathbb{R}^3$.

Proposition 145. *Let $M \in \mathrm{PSL}_2(\mathbb{C})$. Then the following are equivalent:*

1. $M \in \mathrm{PSU}_2(\mathbb{C})$
2. $||M||^2 = 2$
3. *M has the form* $\begin{pmatrix} a & b \\ -\bar{b} & \bar{a} \end{pmatrix}$.
4. *M is elliptic with fixed points of the form α and $-1/\overline{\alpha}$.*

5. M is elliptic with an axis passing through j.

Proof. The equivalence of (2) and (3) comes from the identity

$$|a - \overline{d}|^2 + |b + \overline{c}|^2 = ||M||^2 - 2.$$

That (3) implies (1) is a computation. If M is unitary, then, expanding $M^*M = I$, we obtain from the diagonal entries,

$$a\overline{a} + c\overline{c} = 1 = b\overline{b} + d\overline{d},$$

from which $||M||^2 = 2$ follows. Therefore (1) implies (2).

Next, the geodesics passing through j are exactly those having endpoints of the form α and $-1/\overline{\alpha}$ (an exercise using Pythagorean theorem). This shows (4) and (5) are equivalent.

Next, if M is unitary, then, using the form in (3), we can verify that $(x, y)^t$ is an eigenvector if and only if $(-\overline{y}, \overline{x})^t$ is an eigenvector. We can also verify that M is elliptic using the form of (3), since $a\overline{a} + b\overline{b} = 1$ ipmlies $tr\ (M) = r\Re(a) \leq 2$. This shows (3) implies (4).

Finally, the condition (5) implies M fixes j. Writing the general form of the action of M on H^3 and imposing this condition, we obtain several equations: $ad - bc = 1 = d\overline{d} + c\overline{c}$ and $b\overline{d} + a\overline{c} = 0$. Some brief algebra implies $a = \overline{d}$ and $b = -\overline{c}$, so (5) implies (3). □

Exercise 146. *Choosing a stereographic projection in $\mathbb{R}^3$ that fixes j and takes $\widehat{\mathbb{C}}$ (the xy-plane) to a sphere, the unitary transformations correspond to rotations of that sphere. Therefore* $\mathrm{PSU}_2(\mathbb{C}) \cong SO(3)$.

We have now determined the point stabilizer of a point in the upper half space.

Corollary 147. *The stabilizer of the point j in the upper half space is* $\mathrm{PSU}_2(\mathbb{C})$. *Therefore any point stabilizer for a point in H^3 is conjugate to* $\mathrm{PSU}_2(\mathbb{C})$.

The stabilizer of a point on the boundary requires less work.

Proposition 148. *The stabilizer of ∞ in $\mathrm{PSL}_2(\mathbb{C})$ consists of the upper triangular matrices, i.e.*

$$B = \left\{ \begin{pmatrix} a & b \\ 0 & a^{-1} \end{pmatrix} : a \in \mathbb{C}^*, b \in \mathbb{C} \right\}.$$

The stabilizer of any boundary point is conjugate to this group.

There is an exact sequence

$$1 \longrightarrow \mathbb{C} \longrightarrow B \longrightarrow \mathbb{C}^* \longrightarrow 1$$

where the map $B \to \mathbb{C}^*$ is given by

$$\begin{pmatrix} a & b \\ 0 & a^{-1} \end{pmatrix} \mapsto a^2,$$

or in other words,

$$(z \mapsto \alpha z + b) \mapsto \alpha.$$

The kernel of this map consists of the translations $z \mapsto z + c$, for $c \in \mathbb{C}$, which is isomorphic to $\mathbb{C}$ under addition.

In fact, $B \cong \mathbb{C} \rtimes \mathbb{C}^*$. For, we may compute

$$\begin{pmatrix} a & b \\ 0 & a^{-1} \end{pmatrix} \begin{pmatrix} 1 & c \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a^{-1} & -b \\ 0 & a \end{pmatrix} = \begin{pmatrix} 1 & a^2 c \\ 0 & 1 \end{pmatrix}.$$

In other words, conjugation is a homomorphism,

$$\mathbb{C}^* \to \mathrm{Aut}(\mathbb{C}), \quad \alpha \mapsto (z \mapsto \alpha z).$$

Proposition 149. *The stabilizer of the set $\{0, \infty\}$ in $\mathrm{PSL}_2(\mathbb{C})$ consists of the diagonal and antidiagonal elements, i.e.*

$$\left\{ \begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix} : a \in \mathbb{C}^* \right\} \cup \left\{ \begin{pmatrix} 0 & a \\ -a^{-1} & 0 \end{pmatrix} : a \in \mathbb{C}^* \right\}$$

The stabilizer of any two-element set in $\widehat{\mathbb{C}}$ is conjugate to this group.

5.3 'Small' subgroups of $\mathrm{PSL}_2(\mathbb{C})$

There are some special types of subgroups of $\mathrm{PSL}_2(\mathbb{C})$ that come up as exceptions to theorems, because they are too 'small', or don't move everything around enough.

Definition 150. *Let $\Gamma < \mathrm{PSL}_2(\mathbb{C})$. Then Γ is* reducible *if the group stabilizes some point on ∂H^3. If not, it is* irreducible*.*

In light of the previous section, this is equivalent to Γ being a subset of some conjugate of B.

Definition 151. *Let $\Gamma < \mathrm{PSL}_2(\mathbb{C})$. Then Γ is* elementary *if the orbit of some point of $H^3 \cup \partial H^3$ is finite. If not, it is* non-elementary*.*

For example, reducible groups are elementary.

Proposition 152. *Any abelian group is elementary.*

Proof. Suppose Γ is abelian. We have seen that any pair of commuting elements are either elliptic (inversions, in fact) or they have a common fixed point (Proposition 143). Therefore, ff there are some non-elliptic elements, then all elements share a common fixed point. If they are all elliptic or the identity, then they have a common fixed point by Proposition 144. $\square$

Let $g, h \in \mathrm{PSL}_2(\mathbb{C})$. As a consequence of the characterization of sharing fixed points, $\langle g, h\rangle$ is reducible if and only if $tr\,[g,h] = 2$.

Theorem 153. *The elementary groups of* $\mathrm{PSL}_2(\mathbb{C})$ *come in the following types:*

1. *(Type I) A subgroup of the stabilizer of a point in* H^3*.*
2. *(Type II) A reducible group.*
3. *(Type III) A subgroup of the stabilizer of a two-point set in* $\widehat{\mathbb{C}}$*.*

Proof. Each type is elementary by definition.

Now, suppose Γ is elementary. We will consider four cases.

Case Ia: The finite orbit is not in $\widehat{\mathbb{C}}$. Let $g \in \Gamma$, and let x be in the orbit. Then the elements $g^m(x)$ must eventually repeat. Therefore g^m fixes an element of H^3 and hence it, and indeed g itself, are elliptic (or the identity). Thus Γ consists entirely of elliptic and identity elements, and by Proposition 144, is of Type I.

Case Ib: The finite orbit is of size ≥ 3. Let $g \in \Gamma$, and let x be in the orbit. As before, some power of g fixes x. Taking the least common multiple of the various powers for the various x in the orbit, we discover that some power of g fixes every element of the orbit. Therefore $g^m = id$ for some m, hence g is elliptic, as above, and we are in Type I.

Case II: The finite orbit is a fixed point in $\widehat{\mathbb{C}}$. We are in Type II.

Case III: The finite orbit is a two-element cycle in $\widehat{\mathbb{C}}$. We are in Type III. □

Proposition 154. *Let* Γ *be a non-elementary subgroup of* $\mathrm{PSL}_2(\mathbb{C})$*. Then* Γ *contains infinitely many loxodromic elements in such a way that no two share a fixed points.*

Proof. We follow Beardon, Theorem 5.1.3.

Suppose first that $\Gamma < \mathrm{PSL}_2(\mathbb{C})$ contains no loxodromic elements. We will show that it must be elementary. For, then it consists of the identity, and elliptic and parabolic elements. If there are no parabolics, then it is elementary. Therefore assume it has a parabolic, and conjugate so it is $z \mapsto z+1$. Now, by assumption $tr^{\,2} \in [0,4]$ for all elements of Γ. In particular, if $h : z \mapsto (az+b)/(cz+d) \in \Gamma$, then $tr^2(g^n h) \in [0,4]$. But $tr^{\,2}(g^n h) = (a+d+nc)^2$. Therefore $c = 0$, and Γ fixes ∞, so is elementary.

Therefore any non-elementary $\Gamma < \mathrm{PSL}_2(\mathbb{C})$ contains at least one loxodromic element, say g, fixing some points α and β. Let $h \in \Gamma$ not fix α and β (if this doesn't exist, Γ is once again elementary). Let $f = hgh^{-1}$. Then f is also loxodromic.

First, consider the case that f shares no fixed points with g. Then, for each $n \in \mathbb{Z}$, $r_n : g^n f g^{-1}$ has two fixed points, $g^n h(\alpha)$ and $g^n h(\beta)$, and these never coincide with α nor β. Since g is loxodromic, the sequences $g^n h(\alpha)$ and $g^n h(\beta)$ are non-stabilizing sequences converging to the attracting fixed point of g. Therefore the r_n form an infinite family of loxodromics, no two of which share a point.

Now, instead suppose that f and g share a single fixed point α. Then let $p = [g, f]$. By Proposition 143, g and f do not commute. But by Proposition 142, $tr^{\,2}(p) = 4$, so p is parabolic. It must also fix α. Since Γ is not elementary, there is some $q \in \Gamma$ not fixing α. Therefore $r = qpq^{-1}$ is parabolic, not fixing α. Hence it is fixed-point-disjoint with either f or g; assume it is g. We can replace f with the loxodromic $f' = rgr^{-1}$ sharing no fixed points with g. The proof is then completed as in the last case, with f' and g in place of f and g. □

6 Discontinuous action and fundamental domains

6.1 Discontinuous groups

In this section, we will characterize discrete subgroups of $\mathrm{PSL}_2(\mathbb{C})$ as those which act discontinuously, to be defined below. This allows us to define the notion of a fundamental domain.

Consider the general setting of a topological space X and a group G of homeomorphisms of X.

Definition 155. *The group G acts* discontinuously *if, for every compact set $K \subset X$, $g(K) \cap K = \emptyset$ for all but finitely many $g \in G$.*

Be warned that there are a plethora of equivalent and almost equivalent definitions in the literature. We will stick to this, which agrees with Beardon and Maclachlan and Reid.

Clearly, if G acts discontinuously, so does any subgroup. Furthermore, if G acts discontinuously, then the stabilizer of any point of X is finite.

:discreteorbits **Proposition 156.** *If G acts discontinuously on X, then all orbits G_x for $x \in X$ are discrete.*

Proof. Suppose there is a sequence of distinct elements in the orbit of some x,

$$g_1(x), g_2(x), \ldots \to g_0(x).$$

But then

$$K = \{g_n(x)\}_{n=1}^{\infty} \cup \{g_0(x)\} \cup \{x\}.$$

is a compact set. However, $g(K) \cap K \neq \emptyset$ for any $g = g_n$. Therefore G does not act discontinuously. □

Exercise 157. *Give an example where the converse fails. Now give an example where the point stabilizers are all finite and the converse fails. Now prove the converse, if possible, for a group of isometries on a metric space.*

thm:discdisc **Theorem 158.** *A subgroup of $\mathrm{PSL}_2(\mathbb{R})$ acts discontinuously on H^2 if and only if it is discrete. A subgroup of $\mathrm{PSL}_2(\mathbb{C})$ acts discontinuously on H^3 if and only if it is discrete.*

The proof must make use of the relationship between the topology on $\mathrm{PSL}_2(\mathbb{C})$ and the topology on H^3. This will take the form of the following lemma.

lemma:twotops **Lemma 159.** *Let $g \in \mathrm{PSL}_2(\mathbb{C})$. Then*

$$||g||^2 = 2\cosh\rho(j, g(j)).$$

Proof. A general g, expressed as

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}, \quad ad - bc = 1,$$

acts by

$$g(j) = \frac{(b\overline{d} + a\overline{c}) + j}{|c|^2 + |d|^2}.$$

Then we have,

$$\cosh^2\left(\frac{\rho(j, z + tj)}{2}\right) = \frac{|z|^2 + (t+1)^2}{4t}.$$

Now setting $z + tj = g(j)$, and noting that

$$|b\overline{d} + a\overline{c}|^2 + 1 = (|a|^2 + |b|^2)(|c|^2 + |d|^2),$$

we can simplify to discover that

$$||g||^2 + 2 = 4\cosh^2\left(\frac{\rho(j, z + tj)}{2}\right).$$

Applying a trigonometric identity, we obtain the theorem. $\square$

Proof of Theorem 158. This proof follows Beardon, Theorem 5.3.2. Let $\Gamma < \mathrm{PSL}_2(\mathbb{C})$. We leave the $\mathrm{PSL}_2(\mathbb{R})$ case to the reader.

Suppose Γ is not discrete, so that there exists a non-stabilizing sequence of matrices $g_n \in \Gamma$ approaching I. Then $||g_n||^2 \to 2$. Now consider the sequence $g_n(j) \in H^3$. By Lemma 159, this sequence converges to j. But then the orbit of j is not discrete, hence Γ does not act discontinuously.

Conversely, suppose instead that Γ is discrete. We have seen that the set

$$\{g \in \Gamma : ||g|| < k\}$$

is finite for any k, by Proposition 16. Therefore, by Lemma 159, the set

$$\{g \in \Gamma : \rho(j, g(j)) < k\}$$

is finite for any k. Let B_k be the hyperbolic ball of radius k around j. Now, if $x \in g(B_k) \cap B_k$, then

$$\rho(j, g(j)) \leq \rho(j, x) + \rho(x, g(j)) < 2k.$$

Therefore, for any hyperbolic ball K around j, $g(K) \cap K = \emptyset$ for all but finitely many g.

More generally, let $K \subset H^3$ be compact. Then it is contained in some closed hyperbolic ball B, and if $g(K) \cap K \neq \emptyset$ for some g, then $g(B) \cap B \neq \emptyset$ for that g. Therefore, it suffices to assume K is a closed hyperbolic ball, and we are done. □

Exercise 160. *Prove the case we left out: that $\Gamma < \mathrm{PSL}_2(\mathbb{R})$ is discrete if and only if it acts discontinuously on H^2.*

6.2 Fundamental domains

For this entire section, let $(G, X) = (\mathrm{PSL}_2(\mathbb{C}), H^3)$ or $(\mathrm{PSL}_2(\mathbb{R}), H^2)$.

Definition 161. *Then a* fundamental domain *for the action of $\Gamma < G$ on X is a closed subset $\mathcal{F}$ of X, whose interior is denoted $\mathcal{F}^o$, such that*

1. $\bigcup_{g \in \Gamma} g(\mathcal{F}) = X$.
2. $\mathcal{F}^o \cap g(\mathcal{F}^o) = \emptyset$ *for every non-identity $g \in \Gamma$.*
3. *$\mathcal{F}$ has boundary of measure zero.*

We now define a *Dirichlet region.* The idea is to choose a base point and then collect all the points which are the nearest elements of their orbits.

Definition 162. *Suppose $\alpha \in X$ has trivial stabilizer in $\Gamma < G$. Then define*

$$\mathcal{F}_\alpha = \{\beta \in X : \rho(\alpha, \beta) \leq \rho(\alpha, g(\beta)) \ \forall g \in \Gamma\}.$$

First, let us address the existence of such points α.

prop:countable **Proposition 163.** *Let $\Gamma < G$ act discontinuously. Then Γ is countable.*

Corollary 164. *If $\Gamma < G$ acts discontinuously, then there exist points of X which are not fixed by any $g \in \Gamma$ other than $g = id$.*

Proof of Corollary. **Case I: $G = \mathrm{PSL}_2(\mathbb{R})$, $X = H^2$.** Since Γ is countable, and each non-identity element has at most one fixed point in X, the points fixed by any non-identity $g \in \Gamma$ are countable. But the points of H^2 are uncountable.

Case II: $G = \mathrm{PSL}_2(\mathbb{C})$, $X = H^3$. Each non-identity element's H^3-fixed-points lie in a plane of $\mathbb{R}^3$ intersecting H^3. Since Γ is countable, the union of fixed points are contained in a countable union of planes. But H^3 cannot be covered by a countable union of planes. □

Exercise 165. *Verify the claim used above, that H^3 cannot be covered by a countable union of planes in $\mathbb{R}^3$.*

To prove Proposition 163, we will use the following result, left as an exercise.

Exercise 166. *Show that any discrete subset of $\mathbb{R}^n$ must be countable. Hint: do $\mathbb{R}$ first.*

Proof of Proposition 163. First, suppose that Γ acts discontinuously. Let $x \in H^3$. The stabilizer Γ_x of x must be finite, since any $g \in \Gamma_x$ satisfies $g(K) \cap K \neq \emptyset$ where K is the compact set $\{x\}$. Furthermore, since Γ acts discontinuously, its orbits are discrete. Hence $\Gamma(x)$, as a discrete subset of $\mathbb{R}^3$, is countable. Now, $\Gamma(x)$ is in bijection with the collection of cosets Γ/Γ_x. Since $\Gamma(x)$ and Γ_x are countable, so is Γ. $\square$

This shows that Dirichlet regions exist. We will show that it is a fundamental domain by examining its hyperbolic geometry: we need a little terminology.

Definition 167. *In the hyperbolic plane, the* perpendicular bisector *of a geodesic segment $[\alpha, \beta]$ is the unique geodesic passing through the midpoint of $[\alpha, \beta]$ perpendicularly.*

In hyperbolic 3*-space, the* perpendicular bisecting plane *(or* perpendicular bisector *as before) of a geodesic segment $[\alpha, \beta]$ is the unique geodesic plane passing through the midpoint of $[\alpha, \beta]$ perpendicularly.*

Lemma 168. *The perpendicular bisector of $[\alpha, \beta]$ is given by the equation*

$$\rho(z, \alpha) = \rho(z, \beta)$$

in z (in both the H^2 and H^3 cases).

Proof. In the H^2 case, we may assume without loss of generality that $\alpha = i$ and $\beta = ri$, some $r > 1$. Let s be the positive square root of r. Then $|z|^2 = r$ is clearly a geodesic perpendicular to the imaginary axis, passing through si, the midpoint of i and ri. The equation

$$\rho(z, i) = \rho(z, ri)$$

becomes, for $z = x + yi$,

$$\frac{|z - i|^2}{y} = \frac{|z - r^2 i|^2}{r^2 y}$$

which simplifies to $|z|^2 = r$.

Similarly, in the H^3 case, we may assume $\alpha = j$ and $\beta = tj$. Then $|z|^2 + t^2 = r$ is clearly the perpendicular bisector. The equation

$$\rho(z + tj, j) = \rho(z + tj, rj)$$

becomes

$$\frac{|z|^2 + (t+1)^2}{4t} = \frac{|z|^2 + (t + r^2)^2}{4tr^2}.$$

which is $|z|^2 + t^2 = r$. $\square$

Theorem 169. *1. If $\Gamma < \mathrm{PSL}_2(\mathbb{R})$ is discrete, then any Dirichlet region $\mathcal{F}$ is a convex connected fundamental domain for Γ in H^2. Furthermore, $\mathcal{F}$ is a polygon, its boundary consists of a union of geodesic segments (possibly each infinite in length, possibly infinitely many).*

2. If $\Gamma < \mathrm{PSL}_2(\mathbb{C})$ is discrete, then any Dirichlet region $\mathcal{F}$ is a convex connected fundamental domain for Γ in H^3. Furthermore, $\mathcal{F}$ is a polyhedron, its boundary consists of a union of faces, each a polygon (possibly each with infinitely many sides, possibly with vertices or edges on $\widehat{\mathbb{C}}$, possibly infinitely many) on a geodesic plane.

Proof. Write H^n for H^2 or H^3 as appropriate.

First, by the discontinuous action, for any fixed α and β, the orbit of β is discrete, and therefore the set $\{\rho(\alpha, g(\beta)) : g \in \Gamma\}$ is discrete and so has a minimal element. Therefore $\mathcal{F}_\beta$ intersects every point orbit, so it satisfies property (1) of the definition of a fundamental domain.

Now, we describe the Dirichlet region as an intersection of geodesic half-planes or half-spaces (depending if we work in H^2 or H^3). Write

$$\begin{aligned} H_\alpha(g) &= \{\beta \in H^3 : \rho(\alpha,\beta) \le \rho(\alpha, g(\beta))\} \\ &= \{\beta \in H^3 : \rho(\alpha,\beta) \le \rho(g^{-1}(\alpha), \beta)\}. \end{aligned}$$

Then $\mathcal{F}_\alpha = \bigcap_{g\in\Gamma} H_\alpha(g)$. By the foregoing, each $H_\alpha(g)$ is a half-plane or half-space bounded by a geodesic line or plane which is the perpendicular bisector to the geodesic segment $[\alpha, g^{-1}(\alpha)]$ (this is where we use that α has trivial stabilizer). We will call this boundary

$$\begin{aligned} L_\alpha(g) &= \{\beta \in H^3 : \rho(\alpha,\beta) = \rho(\alpha, g(\beta))\} \\ &= \{\beta \in H^3 : \rho(\alpha,\beta) = \rho(g^{-1}(\alpha), \beta)\}. \end{aligned}$$

Since $\mathcal{F}$ is an intersection of half-planes (or half-spaces), it is closed, connected and convex. Since each $L_\alpha(g)$ has measure zero, the boundary of $\mathcal{F}$ has measure zero.

Next suppose $x, y \in \mathcal{F}^o$, $x \neq y$, are in the same Γ orbit. Then $x = g(y)$ for some g. Since $x \in H_\alpha(g)$, $\rho(x,\alpha) \le \rho(g(x),\alpha) = \rho(y,\alpha)$. Now suppose $\rho(x,\alpha) = \rho(y,\alpha)$. Then $x \in L_\alpha(g^{-1})$, i.e., x is not in $\mathcal{F}^o$. Therefore, if $x \in \mathcal{F}^o$, then $\rho(x,\alpha) < \rho(y,\alpha)$. But similarly, $\rho(y,\alpha) < \rho(x,\alpha)$ and we reach a contradiction. □

Note that the Fuchsian group Γ generated by $z \mapsto z+1$ has a fundamental domain that includes a segment of $\widehat{\mathbb{R}}$ as its boundary at infinity (not its boundary in H^3). We do not consider this to violate the theorem above.

Let us consider the very famous example of $\Gamma = \mathrm{PSL}_2(\mathbb{Z})$. Here we take $\beta = 2i$. Let us compute the Dirichlet region. Consider the elements $T : z \mapsto z+1$ and $S : z \mapsto -1/z$. We compute the geodesic lines

$$
\begin{aligned}
L_{2i}(T) &= \{\beta \in H^2 : \rho(\beta, 2i) = \rho(\beta, 2i-1)\} = -1 + i\mathbb{R}^{\geq 0} \\
L_{2i}(T^{-1}) &= \{\beta \in H^2 : \rho(\beta, 2i) = \rho(\beta, 2i+1)\} = 1 + i\mathbb{R}^{\geq 0} \\
L_{2i}(S) &= \{\beta \in H^2 : \rho(\beta, 2i) = \rho(\beta, i/2)\} = \{\beta \in H^2 : |\beta| = 1\}.
\end{aligned}
$$

Let us now verify that the hyperbolic triangle formed by these three sides, which we will denote $\mathcal{F}$, is in fact the Dirichlet region. Note that

$$\mathcal{F} = \{\beta \in H^2 : |\beta| \geq 1, -1/2 \leq \Re(\beta) \leq 1/2\}.$$

If not, then there is some other $L_{2i}(g)$ intersecting the triangle, which would imply some $\beta \in \mathcal{F}^o$ has $g(\beta) \in \mathcal{F}^o$. Write $g : z \mapsto (az+b)/(cz+d)$, $ad - bc = 1$. Applying the conditions $|\beta| > 1$ and $-1/2 < \Re(\beta) < 1/2$, we obtain

$$|c\beta + d|^2 > (|c|^2 - |d|^2) + |cd|.$$

The right hand side is a positive integer, so $|c\beta + d|^2 > 1$. But then

$$\Im(g(\beta)) = \frac{\Im(\beta)}{|c\beta + d|^2} < \Im(\beta).$$

Repeating the argument with g^{-1} applied to $g(\beta)$ instead of g applied to β, we obtain

$$\Im(\beta) < \Im(g(\beta)).$$

This is a contradiction (this follows Katok, Example A Section 3.2).

Definition 170. *A Kleinian or Fuchsian group* Γ *is called* geometrically finite *if it admits a finite-sided Dirichlet domain.*

Exercise 171. *Note that a Fuchsian group could be considered a Kleinian group via* $\mathrm{PSL}_2(\mathbb{R}) < \mathrm{PSL}_2(\mathbb{C})$. *Does the notion of geometrically finite in both senses coincide?*

Definition 172. *A Kleinian or Fuchsian group* Γ *is* of finite covolume *if it has a fundamental domain of finite hyperbolic volume. It is* cocompact *if it has a compact fundamental domain.*

The volume of the Dirichlet domain for $\mathrm{PSL}_2(\mathbb{Z})$ based at $2i$ is, by the Gauss-Bonnet theorem, given by the angles, which are $\pi/3$ and $\pi/3$, so the volume is $\pi/3$. It has three sides. Therefore $\mathrm{PSL}_2(\mathbb{Z})$ is geometrically finite and of finite covolume. The fundamental region, however, is not compact.

Proposition 173. *Suppose* $\mathcal{F}_1$ *and* $\mathcal{F}_2$ *are fundamental domains for a Kleinian group* $\Gamma < \mathrm{PSL}_2(\mathbb{C})$ *or Fuchsian group* $\Gamma < \mathrm{PSL}_2(\mathbb{R})$. *Then, if* $\mathcal{F}_1$ *has finite volume, so does* $\mathcal{F}_2$ *and the volumes are equal.*

For this reason it is reasonable to talk about the covolume of a Kleinian group.

Proof. We follow Katok, Theorem 3.1.1, and we will write vol for the volume, or area, as appropriate.

We will show that $\mathrm{vol}(\mathcal{F}_1) \geq \mathrm{vol}(\mathcal{F}_2)$. We will consider the translates $\mathcal{F}_1 \cap g(\mathcal{F}_2^o)$ for $g \in \Gamma$; these are disjoint sets contained in $\mathcal{F}_1$. Therefore,

$$\begin{aligned}
\mathrm{vol}(\mathcal{F}_1) &\geq \sum_{g\in\Gamma} \mathrm{vol}(\mathcal{F}_1 \cap g(\mathcal{F}_2^o)) \\
&= \sum_{g\in\Gamma} \mathrm{vol}(g^{-1}(\mathcal{F}_1) \cap \mathcal{F}_2^o) \\
&= \sum_{g\in\Gamma} \mathrm{vol}(g(\mathcal{F}_1) \cap \mathcal{F}_2^o) \\
&\geq \mathrm{vol}(\cup_{g\in\Gamma}(g(\mathcal{F}_1) \cap \mathcal{F}_2^o)) \\
&= \mathrm{vol}(\mathcal{F}_2^o) \\
&= \mathrm{vol}(\mathcal{F}_2).
\end{aligned}$$

The second-to-last step holds since the translates of $\mathcal{F}_1$ cover H^3 (or H^2), so their intersections with $\mathcal{F}_2^o$ cover $\mathcal{F}_2^o$.

Now, repeating the argument, we show $\mathrm{vol}(\mathcal{F}_2) \geq \mathrm{vol}(\mathcal{F}_1)$, or we can observe that the inequalities are actually equalities since fundamental domain boundaries are of measure 0. $\square$

Now we will address the relationship between fundamental domains and subgroups.

Proposition 174. *Suppose Γ and Λ are Kleinian (Fuchsian) groups, and suppose $\Lambda < \Gamma$ with $[\Gamma : \Lambda] = n < \infty$. Suppose $g_1, \ldots, g_n$ forms a system of representatives for the right-cosets of Λ. Suppose $\mathcal{F}$ is a fundamental domain for Γ. Then*

1. *$\mathcal{F}' = \bigcup_{i=1}^n g_i(\mathcal{F})$ is a fundamental domain for Λ,*
2. *if Γ is of finite covolume, then $\mathrm{vol}(\mathcal{F}') = n\,\mathrm{vol}(\mathcal{F})$.*

Proof. Following Katok, Theorem 3.1.2.

Let $\alpha \in H^3$. Then $\alpha = g(\beta)$ for some $\beta \in \mathcal{F}$ and some $g \in \Gamma$. Writing $g = hg_i$ for some $h \in \Lambda$, $i = 1, 2, \ldots, n$, we discover that α is the image of $g_i(\beta) \in \mathcal{F}'$ under $h \in \Lambda$. This verifies the first condition of a fundamental domain for $\mathcal{F}'$: its translates cover H^3.

Now suppose that $\alpha, \beta \in \mathcal{F}'^o$, but that $\beta = g(\alpha)$ for some $g \in \Lambda$. Consider a small ball B around α, small enough to be contained in $\mathcal{F}'^o$. Now apply g: the resulting ball $g(B)$ is a ball around β, hence intersecting some $g_j(\mathcal{F})$ for some j. But then, pulling back, B intersects $g^{-1}g_j(\mathcal{F})$. As $B \subset \mathcal{F}'^o$, we conclude $g^{-1}g_j = g_i$ for some i. But then, as $g \in \Lambda$, they are representatives of the same coset, hence $j = i$. Therefore $g = id$.

The other properties are immediate. $\square$

Proposition 175. *If* Γ *is a Kleinian or Fuchsian group, and* $\mathcal{F}$ *is a Dirichlet domain for* Γ*, then*

1. *for every side* L *of* $\partial\mathcal{F}$*, there is a side* L'*, and a* $g_L \in \Gamma$*, such that* $g_L(L) = L'$*,*
2. $g_{L'} = g_L^{-1}$
3. $(L')' = L$

Proof. Suppose α is the base point of $\mathcal{F}$ as a Dirichlet region. Then the side L is a segment of some $L_\alpha(g)$, for $g \in \Gamma$. Then

$$\begin{aligned} g(L_\alpha(g)) &= g(\{\beta \in H^n : \rho(\alpha, \beta) = \rho(\alpha, g(\beta))\}) \\ &= \{g(\beta) \in H^n : \rho(\alpha, \beta) = \rho(\alpha, g(\beta))\}) \\ &= \big\{\beta \in H^n : \rho(\alpha, g^{-1}(\beta)) = \rho(\alpha, \beta)\big\}) \\ &= L_\alpha(g^{-1}). \end{aligned}$$

Therefore we let $L' = L_\alpha(g^{-1}) \cap \mathcal{F}$ and $g_L = g$, and the first property is verified. The others follow immediately. □

Note that it is possible for $L = L'$ when $g^2 = id$.

Proposition 176. *Dirichlet domains are* locally finite*, meaning, for any compact set* K*,* $K \cap g(\mathcal{F}) \neq \emptyset$ *for only finitely many* $g \in \Gamma$*.*

Proof. Let $\mathcal{F}$ be the Dirichlet domain with basepoint α. We will prove this for closed hyperbolic balls K centred on α, which will imply it in general. Suppose $K \cap g(\mathcal{F}) \neq \emptyset$. Suppose K has centre α and radius r. Then there exists $\beta \in \mathcal{F}$ with $\rho(g(\beta), \alpha) < r$. Then

$$\begin{aligned} \rho(\alpha, g(\alpha)) &\leq \rho(\alpha, g(\beta)) + \rho(g(\beta), g(\alpha)) \\ &\leq \rho(\alpha, g(\beta)) + \rho(\beta, \alpha) \\ &\leq \rho(\alpha, g(\beta)) + \rho(g(\beta), \alpha) \\ &\leq 2r. \end{aligned}$$

However, by discreteness, only finitely many g can satisfy this. □

Proposition 177. *Any Kleinian (Fuchsian)* Γ *is generated by the side-pairing transformations for any Dirichlet domain* $\mathcal{F}$*.*

Proof. Let Λ be the subgroup of Γ generated by

$$\{g \in \Gamma : g(\mathcal{F}) \cap \mathcal{F} \neq \emptyset\}.$$

Our first step will be to show that $\Lambda = \Gamma$. For some $\beta \in H^n$, suppose $g(\beta) \in \mathcal{F}$ and $h(\beta) \in \mathcal{F}$. Then also $h(\beta) \in hg^{-1}(\mathcal{F})$. This implies $gh^{-1} \in \Lambda$. So in particular, we have a well-defined map

$$\phi : H^n \to \Gamma/\Lambda, \quad \beta \mapsto \Lambda g, \text{ for any } g \text{ such that } \beta \in g(\mathcal{F}).$$

The task is to show that ϕ is a constant map. Since $\phi(\beta) = \Lambda$ for $\beta \in \mathcal{F}$, this would imply all cosets of Λ are trivial, i.e. $\Lambda = \Gamma$.

For this, suppose $\beta \in H^n$. First, $\beta \in g_i(\mathcal{F})$ for only finitely many $g_1, \ldots, g_n$ (by local finiteness).

Now, take

$$S = \bigcup_{i=1}^{n} g_i(\mathcal{F}).$$

Then there is an open neighbourhood U of β contained in S, i.e. $\beta \in U \subset S$. This is possible because, first, the domain satisfies local finiteness, so that the closure of a small ball around β intersects only finitely many translates, and second, because $\mathcal{F}$ and its translates are closed, so we may shrink U to avoid any translate not containing β. Now suppose $\gamma \in U$. Then, $\gamma \in g_i(\mathcal{F})$ for some $1 \leq i \leq n$. Also, $\beta \in g_i(\mathcal{F})$. Therefore

$$\phi(\gamma) = \Lambda g_i^{-1} = \phi(\beta),$$

and ϕ is constant on the open neighbourhood U.

Any function which is constant on some open neighbourhood of each point is constant on the domain, as needed.

This shows that $\Lambda = \Gamma$, i.e. Γ is generated by the elements $g \in \Gamma$ such that $g(\mathcal{F}) \cap \mathcal{F} \neq \emptyset$.

The next task is to show that these elements are all generated by the side-pairing elements. Let $g \in \Gamma$ be such that $g(\mathcal{F}) \cap \mathcal{F} \neq \emptyset$. Then we must show g is generated by side-pairing elements. Let $\beta \in g(\mathcal{F}) \cap \mathcal{F}$. As in the first part of the proof, it is the case that $\beta \in g_i(\mathcal{F})$ for only finitely many $g_1, \ldots, g_n$, and we may take $g_1 = id$, $g_2 = g$ and let

$$S = \bigcup_{i=1}^{n} g_i(\mathcal{F}).$$

Then there is an open neighbourhood U of β contained in S, i.e. $\beta \in U \subset S$. By shrinking U, using local finiteness, we may assume it intersects no $L_\alpha(g)$ not actually containing β (here, α is the basepoint of the Dirichlet domain).

Now, let $\gamma_0 \in U \cap \mathcal{F}$ and $\gamma_1 \in U \cap g(\mathcal{F})$. Then there is a path in p joining γ_0 to γ_1 and not passing through any intersections of the $L_\alpha(g)$, i.e. only passing through a single $L_\alpha(g)$ at once. To see this, note that the points to be avoided form a codimension > 1 set, so can successfully be avoided in a ball.

Travelling along the path, we see a sequence of $g_i(\mathcal{F})$: we can label this sequence of the g_i's as follows:

$$h_0 = id, h_1, h_2, \ldots, h_m = g.$$

Then, in particular, $h_i(\mathcal{F}) \cap h_{i+1}(\mathcal{F}) \cap U$ is a segment of an $L_\alpha(h_{i+1}^{-1}h_i)$. In other words, $h_{i+1}(\mathcal{F})$ shares a side with $h_i(\mathcal{F})$, and $h := h_{i+1}^{-1}h_i$ is a side pairing transformation. But then

$$h_{i+1} = h_i h^{-1}.$$

As $h_0 = id$, this implies recursively that all the h_i, including g itself, are generated by side-pairing transformations. □

Corollary 178. *Any geometrically finite Kleinian or Fuchsian group is finitely generated.*

Proposition 179. *Any Kleinian or Fuchsian group with a compact Dirichlet domain is geometrically finite.*

Proof. Let $\mathcal{F}_\alpha$ be a Dirichlet domain. Consider a closed hyperbolic ball centred on α. As it expands, it will eventually touch every side of the domain, by compactness (eventually it will contain the whole domain).

Expanding slightly more, it now intersects its own translates by each side-pairing element. If there are infinitely many sides, the it intersects infinitely many of its own translates, a contradiction to the discontinuous action. Therefore the domain has finitely many sides. □

More generally, actually, we have:

Proposition 180 (Garland-Raghunathan-1970 – Wielenberg-1977)**.** *Any Kleinian group of finite covolume is geometrically finite.*

Proof. You will find a brief discussion in Lemma 3.6.4 of Marden, or Theorem 2.7 of Elstrodt, Grunewald and Mennicke. The proof requires some tools we don't have yet. □

Corollary 181. *Any finite covolume Kleinian or Fuchsian group is finitely generated.*

6.3 Limit sets and regular sets

We turn our attention to the action of a discrete group on the boundary $\widehat{\mathbb{C}}$. Here it need not act discontinuously.

Definition 182. *A point $z \in \widehat{\mathbb{C}}$ is a* limit point *for the action of Γ if there is a point $w \in \widehat{\mathbb{C}}$ and a sequence of distinct elements $g_1, g_2, \ldots \in \Gamma$ such that $g_i(w) \to z$.*

The collection of limit points is called the limit set *of Γ and is denoted $\Lambda(\Gamma)$.*

Proposition 183. *The limit set contains any point fixed by a loxodromic or parabolic element of Γ.*

Proof. Suppose $g \in \Gamma$ is loxodromic or parabolic. Let z be a fixed point and w a non fixed point. Then z is an attracting fixed point for g or g^{-1}; assume it is the former. Then $g^n(w) \to z$. □

Corollary 184. *If $\Lambda(\Gamma)$ is finite, then Γ is elementary or trivial.*

Proof. We showed that any non-trivial non-elementary group contains infinitely many loxodromic elements in such a way that no two share a common fixed point. Therefore the union of their fixed points must be infinite. □

Now, it will be helpful to give a variety of equivalent definitions of the limit set. One of the most important is as the accumulation points of H^3 orbits. These orbits can be considered in $H^3 \cup \partial H^3$, which is compact. So infinite orbits do have accumulation points, but they cannot be in H^3 by discreteness: they must lie on the boundary.

thm:equivlimits **Theorem 185.** *Let Γ be a non-elementary Kleinian group. Then the following sets are all equal:*

item:limitset 1. *The limit set of Γ*

item:acc-chat 2. *The accumulation points of the orbit of any $z \in \widehat{\mathbb{C}}$.*

tem:lox-closure 3. *The closure of the set of loxodromic fixed points.*

item:min-inv 4. *The minimal non-empty closed subset of $\widehat{\mathbb{C}}$ invariant under Γ.*

item:acc-h3 5. *The accumulation points of $\Gamma(\alpha)$ for any $\alpha \in H^3$.*

We will need to make use of a lemma that provides a type of uniform convergence. This lemma can also be viewed as an explicit version of the Lipshitz condition for the rational maps and the chordal metric. See the exercise following.

a:chordalshrink **Lemma 186.** *Suppose that K is a compact subset of some open set $D \subset \widehat{\mathbb{C}}$. Then there is a constant C, depending only on D and K, so that for any g not taking values 0 or ∞ on D, we have*

$$d(gz, gw) \leq \frac{C}{||g||^2} d(z,w), \quad \forall\, z, w \in K,$$

where d represents the chordal metric.

Proof. This follows Beardon, Theorem 4.5.5. Let $g : z \mapsto \frac{az+b}{cz+d}$, $ad - bc = 1$. Consider the infimum

$$m := \inf\{d(z,w) : z \in K, w \notin D\}.$$

By assumption, $g^{-1}(\infty) \notin D$. Therefore for any $z \in K$ we have

$$\begin{aligned} m &\leq d(z, g^{-1}(\infty)) \\ &\leq \frac{2|cz+d|}{(1+|z|^2)^{1/2}(|c|^2+|d|^2)^{1/2}}. \end{aligned}$$

Similarly,

$$\begin{aligned} m &\leq d(z, g^{-1}(0)) \\ &\leq \frac{2|az+b|}{(1+|z|^2)^{1/2}(|a|^2+|b|^2)^{1/2}}. \end{aligned}$$

From these two, we may conclude that

$$m^2(1+|z|^2)||g||^2 \leq 2|az+b|^2 + 2|cz+d|^2.$$

Using the chordal metric we have,

$$\begin{aligned} \frac{d(g(z),g(w))^2}{d(z,w)^2} &\leq \frac{(1+|z|^2)(1+|w|^2)}{(|az+b|^2+|cz+d|^2)(|aw+b|^2+|cw+d|^2)} \\ &\leq \frac{4}{m^4||g||^4}. \end{aligned}$$

□

Exercise 187. *Let $\phi : \mathbb{P}^1(\mathbb{C}) \to \mathbb{P}^1(\mathbb{C})$ be a rational map. Show that ϕ is Lipschitz with respect to the chordal metric, i.e. there exists a constant C, depending only on ϕ, such that for all $\alpha, \beta \in \mathbb{P}^1(\mathbb{C})$,*

$$d(\phi(\alpha), \phi(\beta)) \leq C d(\alpha, \beta).$$

Furthermore, find the explicit constant for a Möbius transformation ϕ.

lemma:closed **Lemma 188.** *The accumulation points of a single orbit in $\widehat{\mathbb{C}}$ form a closed set.*

Proof. Suppose z_m are accumulation points of the orbit of z, i.e.

$$g_{n,m}(z) \to z_m, \text{ as } n \to \infty.$$

Suppose also that $z_m \to z_0$ as $m \to \infty$. We wish to find a sequence of h_k such that $h_k(z) \to z_0$ as $k \to \infty$. We have

$$d(g_{n,m}(z), z_0) \leq d(g_{n,m}(z), z_m) + d(z_m, z_0).$$

For a given ϵ, choose m large enough so the second part is less than $\epsilon/2$, and then, contingent on m, choose n large enough so that the first part is less than $\epsilon/2$. For $\epsilon = 1/k$, let $h_k = g_{n,m}$. Then

$$d(h_k(z), z_0) < 1/k,$$

and so the sequence converges. □

emma:shrinkloxo **Lemma 189.** *If D is an open disk and $g(\overline{D}) \subset D$ for some Möbius transformation g, then g is loxodromic with a fixed point in D.*

Proof. By conjugation, suppose $g(\infty) = \infty$. Then D is a Euclidean circle in $\widehat{\mathbb{C}}$ as it cannot pass through ∞. If g is identity, elliptic or parabolic fixing ∞, then it acts on $\mathbb{C}$ as a Euclidean isometry, so it cannot take $g(\overline{D}) \subset D$. Therefore g is loxodromic. As any non-fixed point approaches the attracting fixed point, that attracting fixed point must lie in D. □

We also need a geometric tool here: for any $\alpha \in H^3$ and $D \subset \widehat{\mathbb{C}}$, the *visual angle* $\theta(\alpha, D)$ is the supremum of the angles at α between geodesics joining α to points in D. If $g \in \mathrm{PSL}_2(\mathbb{C})$ then conformality guarantees that the visual angle is preserved, i.e.

$$\theta(g(\alpha), g(D)) = \theta(\alpha, D).$$

Proof of Theorem 185. (1) **contains each** (2)**.** This is immediate.

Each (2) **contains** (3)**.** Suppose w is a loxodromic fixed point, say of a loxodromic g. Then w is in the accumulation set of any s not fixed by g since $g^n(s)$ or $g^{-n}(s)$ approaches w. Suppose s is the other fixed point of g. Since Γ is non-elementary, the orbit of s includes some other point not fixed by g and so, by the previous case, w is in the accumulation set of s. Therefore the accumulation set of any point in $\widehat{\mathbb{C}}$ is closed (by Lemma 188) and includes the loxodromic fixed points.

(3) **contains** (1)**.** We will show that no accumulation point of an orbit can be in the complement of the closure of the loxodromic fixed points. The proof follows Beardon, Theorem 5.3.9.

First, conjugate so that 0 and ∞ are loxodromic fixed points. Let D be the complement of the closure of the loxodromic fixed points.

Now, suppose, for a contradiction, that some $g_n(z) \to w$, where $w \in D$. Then there is a compact set K containing w in its interior, and also containing z, so that $K \subset D$. Now, if $g_n(y) = 0$ or ∞ for some y in D, then y is a loxodromic fixed point, a contradiction. So we may now apply Lemma 186. We find that, for all $z_1, z_2 \in K$,

$$d(g_n(z_1), g_n(z_2)) \leq \frac{C}{||g_n||^2}$$

for some constant C.

Therefore, for sufficiently large n, $g_n(L) \subset L^o$ for any closed disk L inside K, around w. Lemma 189 implies g_n is loxodromic with a fixed point inside L, which is a contradiction.

(4) **exists and is equal to** (1)**.** Any non-empty closed Γ-invariant set contains a point. By Γ-invariance and closure, it contains all accumulation points of the orbit of that point. Hence it contains at least one (2), but we have seen that every (2), which is closed and Γ-invariant, is equal to (1).

Each (5) contains (3). Suppose z is in the closure of the set of loxodromic fixed points. The loxodromic fixed points are evidently accumulation points of $\Gamma(\alpha)$, and an accumulation point of accumulation points is an accumulation point (as in the proof of Lemma 188).

(3) **contains each (5).** (This follows Matsuzaki and Taniguchi, Section 2.1.) Suppose $z \in \widehat{\mathbb{C}}$ is an accumulation point of $\Gamma(\alpha)$, say $g_n(\alpha) \to z$. Suppose z is not a loxodromic fixed point. We wish to find it as an accumulation point of loxodromic fixed points. Consider the sequence $g_n^{-1}(\alpha)$. We can pass to a subsequence to guarantee this converges to some $w \in \widehat{\mathbb{C}}$ (by compactness).

Now choose a loxodromic not fixing z or w, but with some fixed point $t \in \widehat{\mathbb{C}}$ (which we can do since Γ is non-elementary). The orbit of a loxodromic fixed point consists entirely of loxodromic fixed points. We will show that the orbit of t accumulates at z.

Take an open disc D around w in $\widehat{\mathbb{C}}$, not containing t or z. Since $g_n^{-1}(\alpha) \to w$, we have

$$\theta(\alpha, g_n(D)) = \theta(g_n^{-1}(\alpha), D) \to 2\pi.$$

That implies that

$$\theta(\alpha, g_n(\widehat{\mathbb{C}} \setminus D)) \to 0.$$

But $g_n(\widehat{\mathbb{C}} \setminus D)$ must therefore converge to a point. This point must be z since $g_n(\alpha) \to z \in \widehat{\mathbb{C}} \setminus D$. So in particular, z is an accumulation point of the orbit of t, since $t \in \widehat{\mathbb{C}} \setminus D$. $\square$

An immediate consequence of this proof is the following:

Proposition 190. *Let Γ be non-elementary. The orbit of a limit point is dense in the limit set. Furthermore, the limit set is perfect (i.e. closed with no isolated points).*

Proof. The first statement is an immediate consequence of the last proof. Consider a loxodromic fixed point z. Suppose z is fixed by a loxodromic g. Take some other point w not fixed by g. Then $\{g^n(w)\}$ accumulates at z, but is never equal to it. Therefore z is not isolated. If a set has no isolated points, neither does its closure. $\square$

Exercise 191. *Perfect sets in $\mathbb{R}^n$ are uncountable.*

Proposition 192. *Any non-elementary limit set which is not all of $\widehat{\mathbb{C}}$ has empty interior.*

(It's actually true for elementary limit sets too.)

Proof. As $\Lambda(\Gamma)$ is closed, if it is not all of $\widehat{\mathbb{C}}$, then we may take some open set U in its complement. The orbit of this open set is dense in $\Lambda(\Gamma)$, and yet remains in the complement. Therefore $\Lambda(\Gamma)$ can have no interior. $\square$

Proposition 193. *Suppose $\Gamma' < \Gamma$, where Γ' is non-elementary. If the index is finite or Γ' is normal, then $\Lambda(\Gamma') = \Lambda(\Gamma)$.*

Proof. Suppose Γ' has finite index. Then any loxodromic $g \in \Gamma$ has some power in Γ'. Therefore the loxodromic fixed points for Γ and Γ' are the same, so the limit sets are the same.

Suppose instead that Γ' is normal. Then $g\Gamma' g^{-1} = \Gamma'$ for any $g \in \Gamma$. For this reason, if z is a loxodromic fixed point for some $g' \in \Gamma'$, then $g(z)$ is a loxodromic fixed point for the element $gg'g^{-1} \in \Gamma'$. Therefore the orbit of z is contained in $\Lambda(\Gamma')$. But it is dense in $\Lambda(\Gamma)$, so $\Lambda(\Gamma')$, being closed, contains $\Lambda(\Gamma)$. □

Exercise 194. *Classify the finite limit sets.*

Definition 195. *Let Γ be a Kleinian group. Then the* ordinary set*,* regular set *or* set of discontinuity *for Γ is the maximal open set in $\widehat{\mathbb{C}}$ on which Γ acts discontinuously. It is denoted $\Omega(\Gamma)$.*

Proposition 196. *The limit set and ordinary set are complements in $\widehat{\mathbb{C}}$.*

Proof. Following Beardon, Theorem 5.3.10. We will show that the complement of Λ is the maximal open set of discontinuity. First, we show that Γ does act discontinuously away from Λ. Suppose it does not. Then there is some compact subset K such that $K \cap \Lambda = \emptyset$, so that $g_n(K) \cap K \neq \emptyset$ for infinitely many g_n. In other words, there are points $z_n \in K$ with $g_n(z_n) \in K$. Since K is compact, some subsequence of $g_n(z_n)$ converges to some $w \in K$. Restrict to this subsequence. Then $w \notin \Lambda$.

Now, we use Lemma 186, on K. We may conjugate so that 0 and ∞ are in Λ. Then g_n do not take values 0 or ∞ on K. Then,

$$d(g_n(z_n), g_n(z_1)) \leq \frac{C}{||g_n||^2} d(z_n, z_1).$$

Therefore, $g_n(z_1) \to w$, and so $w \in \Lambda$, for a contradiction.

Next, we suppose that Γ acts discontinuously on some open U. By discontinuous action, U cannot contain any loxodromic fixed points, so U cannot intersect Λ. This completes the proof. □

7 Hyperbolic Manifolds and Orbifolds

7.1 Hyperbolic manifolds and fundamental groups

Definition 197. *A* hyperbolic n-manifold *is an n-manifold whose Riemannian metric is hyperbolic, i.e. any open neighbourhood is isometric to an open subset of hyperbolic n-space.*

Theorem 198. *If Γ is a torsion-free Kleinian group, then H^3/Γ is an orientable hyperbolic 3-manifold. Conversely, any orientable hyperbolic 3-manifold is isometric to H^3/Γ for some torsion-free Kleinian group Γ.*

Proof. Since Γ is Kleinian, it acts discontinuously on H^3. Since Γ is discrete, point stabilizers are finite. Since Γ is torsion-free, then, point stabilizers are trivial, and Γ acts freely on H^3. Thus the quotient is an orientable hyperbolic 3-manifold.

Conversely, any orientable hyperbolic 3-manifold M has a unique universal cover up to isometry: that is H^3. The fundamental group $\pi_1(M)$ acts on the cover by deck transformations which are orientation preserving isometries, i.e. $\pi_1(M) \cong \Gamma < \mathrm{PSL}_2(\mathbb{C})$. It must act discontinuously, so that Γ is a Kleinian group. And it must not have torsion. □

In particular, the volume of H^3/Γ is the volume of the fundamental domain.

The same result holds for Fuchsian groups, i.e.

Theorem 199. *If Γ is a torsion-free Fuchsian group, then H^2/Γ is an orientable hyperbolic 2-manifold. Conversely, any orientable hyperbolic 2-manifold is isometric to H^2/Γ for some torsion-free Fuchsian group Γ.*

The proof is the same.

For more background, see the following famous lecture of McMullen: https://vimeo.com/30775825 (I plan to have us view it in class), and accompanying article, *The evolution of geometric structures on 3-manifolds*: http://www.ams.org/journals/bull/2011-48-02/S0273-0979-2011-01329-5/.

7.2 Dirichlet domains, Cayley graphs, and the fundamental group

Definition 200. *Let G be a group and S be a finite generating set of G, not containing the identity. Then the Cayley graph $\mathcal{G}$ of G with respect to S is the graph whose vertices are the elements of G, and which has a directed edge joining g to h if and only if $h = sg$ for some $s \in S$.*

If S consists of elements of order two, then the Cayley graph can be considered an undirected graph.

Now, given a tesselation of H^2 or H^3 by Dirichlet fundamental domains for Γ with respect to α, we can construct the *dual graph* as follows: the vertices are all translates of α, and the edges are geodesic segments $[g(\alpha), hg(\alpha)]$ where h ranges over the side-pairing elements. Therefore, this graph is an embedding of the Cayley graph (with directions forgotten) of Γ with respect to its generating set of side-pairing elements.

This dual graph, as a subset of H^n is invariant under Γ. Therefore it has a well-defined image under the quotient $H^3 \to H^3/\Gamma$ (or $H^2 \to H^2/\Gamma$). Each edge maps to a loop, i.e. an element of the fundamental group based at the image of α. In fact, these loops generate the fundamental group. Furthermore, the 2-cells (minimal cycles) involving α generate the relations in the fundamental group.

7.3 Ends

The next task is to describe the 'ends' of a hyperbolic 3-manifold M.

Definition 201. *Let M be a non-compact hyperbolic 3-manifold. Let $K_1 \subset K_2 \subset \cdots$ be an increasing sequence of compact sets the union of whose interiors covers M. Then let $U_1 \supset U_2 \supset \cdots$ be a decreasing sequence of open sets so that U_n is a connected component of $M \setminus K_n$. Such a sequence is called an* end *of the manifold.*

In other words, the manifolds we are examining may 'extend to infinity' in various directions; each such direction is an 'end'. Our main concern will be with cusp ends: the points on $\widehat{\mathbb{C}}$ where sides of the fundamental domain funnel down to a point. But an end could be of various shapes.

In order to do this, we must examine the stabilizer of a point of the boundary.

7.4 Elementary discrete stabilizers

Proposition 202. *Any discrete elementary group Γ fixing an element of H^3 is finite and isomorphic to the symmetry group of a platonic solid, a finite cyclic group, or a dihedral groups.*

Recall that the symmetry groups of the tetrahedron, octahedron, and icosahedron, are A_4, S_4 and A_5 respectively.

Proof. Assume by conjugation that the common fixed point is j (this preserves discreteness). Then Γ is a subgroup of $\mathrm{PSU}_2(\mathbb{C})$. Since $||g||^2 = 2$ for all $g \in \mathrm{PSU}_2(\mathbb{C})$, and Γ is discrete, Γ must be finite. Recall that $\mathrm{PSU}_2(\mathbb{C}) \cong SO(3)$. But the finite subgroups of $SO(3)$ are exactly the symmetry groups of the platonicsolids, the finite cyclic groups, and the dihedral groups (there are many references for this fact; see also Beardon, Section 5.1, for a direct approach). $\square$

Lemma 203. *Suppose f and g have exactly one fixed point in common, and g is loxodromic. Then $\langle f, g\rangle$ is not discrete.*

Proof. Conjugate so that the common fixed point is ∞. We may also assume without loss of generality that $g : z \mapsto az$, $|a| \neq 1$, and that $f : z \mapsto cz + d$, where $d \neq 0$. Suppose first that $|a| > 1$. Then consider the sequence of maps

$$g^{-n} f g^n : z \mapsto cz + da^{-n}, \quad n = 1, 2, 3, \ldots.$$

Then $g^{-n} f g^n \to (z \mapsto cz)$ is a non-stabilizing convergent sequence, and any subgroup of $\mathrm{PSL}_2(\mathbb{C})$ containing f and g fails to be discrete. If $|a| < 1$, then consider instead $g^n f g^{-n}$ to the same effect. $\square$

prop:discB **Proposition 204.** *Any discrete subgroup of B containing a parabolic element is of the form*

$$\{z \mapsto \alpha z + b : \alpha \in \mu_n, b \in \Lambda\}.$$

where μ_n is the finite group of n-th roots of unity for some $n \in \{1, 2, 3, 4, 6\}$, and Λ is a non-trivial lattice of $\mathbb{C}$ invariant under multiplication by elements of μ_n. Any discrete subgroup of B without parabolic elements is isomorphic to a discrete subgroup of $\mathbb{C}^*$.

Proof of Proposition 204. We have seen that there is a split exact sequence

$$1 \longrightarrow \mathbb{C} \longrightarrow B \longrightarrow \mathbb{C}^* \longrightarrow 1\ .$$

Now let us restrict this sequence to Γ:

$$1 \longrightarrow \Lambda \longrightarrow \Gamma \longrightarrow S \longrightarrow 1\ ,$$

where S is the image of Γ in $\mathbb{C}^*$, and Λ are the translations in Γ.

Suppose Γ contains at least one parabolic element, a translation $t : z \mapsto z + r$ in Λ. By the lemma, Γ contains no loxodromic elements, so any other element is of the form $g : z \mapsto az + b$, where $|a| = 1$. So in particular, S is a subset of the unit circle. Notice that

$$(gtg^{-1} : z \mapsto z + ar) \in \Lambda.$$

By discreteness, the values of a cannot accumulate. So S is a discrete subgroup of the unit circle. The only discrete subgroups of S are finite cyclic groups. Therefore $S = \mu_n$ for some n.

Now, consider the set Λ, which must be a non-trivial discrete subgroup of $\mathbb{C}$, and, as such, a lattice of rank 1 or 2. Furthermore, by the semi-direct product structure of B, $\alpha\Lambda \subset \Lambda$ for any $\alpha \in S$.

There are no lattices invariant under μ_n unless $n \in \{1, 2, 3, 4, 6\}$.

Now, suppose instead that Γ contains no parabolics, so Λ is trivial. Then $\Gamma \cong S$, i.e. Γ is isomorphic to some discrete subgroup of $\mathbb{C}^*$, and therefore cyclic.

This completes the proof. $\square$

Exercise 205. *Verify that the only discrete subgroups of the unit circle are finite cyclic. Verify that the only non-trivial discrete subgroups of $\mathbb{C}$ are lattices of rank 1 or 2.*

Exercise 206. *Verify that there are no lattices invariant under μ_n unless $n \in \{1, 2, 3, 4, 6\}$. Show that the only such lattices when $n \neq 1$ are homothetic to sublattices of the Gaussian or Eisenstein integers.*

Corollary 207. *Any discrete subgroup of B is either*

1. *finite cyclic*

2. *of the form $A \times C$ where A is infinite cyclic generated by a loxodromic, and C is finite*

3. *of the form $A \rtimes C$ where A is infinite cyclic generated by a parabolic, and C is finite*

4. *of the form $A \rtimes C$ where $A \cong \mathbb{Z}^2$ is generated by parabolics, and C is finite*

To verify this corollary, we must show the following.

Exercise 208. *All discrete subgroups of $\mathbb{C}^*$ are isomorphic to groups of type (1) and (2) above.*

Points whose point stabilizers are of the largest possible rank are important.

Definition 209. *Let Γ be a Kleinian group. A point $\alpha \in \widehat{\mathbb{C}}$ is a* cusp *of Γ if the stabilizer of α in Γ contains a free abelian group of rank two.*

Definition 210. *A* horoball at ∞ *is a set of the form*

$$H_\infty(t_0) = \{z + tj : t > t_0\}.$$

A horoball at α *where $\alpha \in \widehat{\mathbb{C}}$ is the image of any $H_\infty(t_0)$ under an isometry taking $\infty \mapsto \alpha$.*

Proposition 211. *Horoballs at α are Euclidean balls in $\mathbb{R}^3$ tangent to $\widehat{\mathbb{C}}$ at α.*

Exercise 212. *Prove the proposition.*

Proposition 213. *Any discrete subgroup of B with parabolics acts as Euclidean isometries on a horoball at ∞.*

Proof. B consists of maps of the form $z \mapsto az + b$. On H^3, they act as

$$z + tj \mapsto az + b + |a|tj.$$

However, the subgroup being discrete, we see that $|a| = 1$, so it acts on each horoball at ∞, which is a copy of $\mathbb{C}$, as Euclidean similarities on $\mathbb{C}$. □

In other words, the quotient of the horoball by the action of a discrete subgroup Γ of B is a torus, in the case there's a rank two group generated by parabolics (a lattice of translations). In this case the end is isometric to $S^1 \times S^1 \times [0, \infty)$.

Proposition 214. *If a Kleinian Γ contains a parabolic, then it is not co-compact.*

Proof. Proof from Jones and Singerman, *Complex Functions*, Theorem 5.9.9. If Γ is Kleinian with compact fundamental domain $\mathcal{F}$, then the number of g such that $\mathcal{F} \cap g(\mathcal{F}) \neq \emptyset$ is finite. Therefore the following infimum exists and is positive for any $z \in \mathcal{F}$:

$$\eta(z) = \inf\{\rho(z, g(z)) : g \text{ non-elliptic, non-identity}\}.$$

In fact, $\eta(z)$ is a continuous function of z. Therefore it, too, attains its infimum, as z varies in the compact set $\mathcal{F}$; this is also the infimum as z varies in the whole of H^3. But it does not attain 0.

By conjugation, we can assume our parabolic is $z \mapsto z+1$. But we have

$$\rho(z, z+1) \to 0$$

as the imaginary part of z approaches ∞. This contradicts the assertion about the infimum of $\eta(z)$. $\square$

The converse also holds for finite covolume groups.

Proposition 215. *A Kleinian group Γ of finite covolume is co-compact if and only if it contains no parabolic elements.*

We also have the following:

Proposition 216. *Suppose Γ is a Kleinian of finite covolume. Then the cusps occur exactly at the parabolic fixed points, and there are finitely many Γ-equivalence classes of cusps. If Γ is torsion-free, these classes are in bijection with the ends of the corresponding hyperbolic 3-manifold.*

I won't give full proofs of these facts, which can be explained by use of the *thick-thin decomposition* and the 'Margulis lemma'. To explain briefly, there's a notion of *injectivity radius*, the size of the largest ball you can embed in the manifold at a point. The ϵ-thin part is the part where the injectivity radius is bounded above by ϵ. Margulis' result implies that there is a universal radius so that all parts ϵ-thin for ϵ below that radius look like cusp cylinders or cusp tori, i.e. ends approaching parabolic fixed points (former with stabilizer of parabolic rank 1, latter with stabilizer of parabolic rank 2); see the pictures I drew in class. If the parabolic rank is 1, then the resulting end has infinite volume.

The following related statement about fundamental domains near a parabolic fixed point is relevant.

hm:prec-inv-hor **Theorem 217.** *Suppose Γ is Kleinian with parabolic fixed point. Conjugate so the relevant parabolic is $z \mapsto z+1$. Then the open horoball at infinity $H = \{z + tj : t > 1\}$ is precisely invariant under the stabilizer G of ∞, i.e.*

1. *$g(H) = H$ for any $g \in G$*
2. *$g(H) \cap H = \emptyset$ for any $g \in \Gamma \setminus G$*

In particular, if you zoom in to the parabolic fixed point, you eventually see only the part of the Kleinian group that is generated by the parabolic elements with that fixed point.

The proof uses a result of independent interest that we won't prove in these notes, but is a most useful necessary condition for discrete non-elementary groups (often used to prove non-discreteness).

prop:jorg **Proposition 218** (Jorgensen's Inequality). *If f and g generate a discrete non-elementary group, then*

$$|tr^2(f) - 4| + |tr(fgf^{-1}g^{-1} - 2| \geq 1.$$

This lower bound is best possible; see Beardon, Theorem 5.4.1.

Exercise 219. *It is an application of Jorgensen's Inequality that a non-elementary subgroup of* $\mathrm{PSL}_2(\mathbb{C})$ *is discrete if and only if every subgroup generated by two elements is discrete. (Soln: see Theorem 5.4.2 in Beardon. Of course, one direction of this is obvious. Hint: Suppose G is not discrete and suppose a sequence of elements converging to the identity. Generate groups with these paired with various loxodromics of differing fixed points.)*

Proof of Theorem 217. We have already observed the first property. We need to show that $g(H) \cap H = \emptyset$ whenever g does not fix ∞.

Let $g \in \mathrm{PSL}_2(\mathbb{C})$ has the form $g : z \mapsto (az + b)/(cz + d)$. Suppose it does not fix ∞, so $c \neq 0$.

Now, Γ contains $f : z \mapsto z + 1$. Suppose it also contains g as above. Then, as Γ is discrete and non-elementary, by Jorgensen's Inequality (Proposition 218), we have

$$|tr^2 f - 4| + |tr(fgf^{-1}g^{-1}) - 2| \geq 1.$$

Since $tr\ ^2 f = 4$ and $tr\ (fgf^{-1}g^{-1} = 2 + c^2$, we conclude that $|c| \geq 1$.

Now, using the explicit form (1) of the action on H^3 of g, if $|c| \geq 1$ and $t \geq 1$, then $g(z + tj)$ has height

$$\frac{t}{|cz + d|^2 + |c|^2 t^2} \leq \frac{1}{|c|^2 t} \leq 1.$$

Therefore the horoball H is taken entirely outside itself. □

7.5 Orbifolds and Selberg's Lemma

Definition 220. *A* hyperbolic 3-orbifold *is a quotient of the form H^3/Γ where Γ is a Kleinian group. The points with non-trivial stabilizer are called* orbifold points *or* singular points.

An orbifold is a manifold if Γ is torsion-free. Our goal in this section will be to prove Selberg's Lemma (following Matsuzaki and Taniguchi, Section 2.3).

theorem:selberg **Theorem 221** (Selberg's Lemma). *Any finitely-generated subgroup Γ of* $\mathrm{SL}_2(\mathbb{C})$ *has a finite-index subgroup which is torsion-free and normal.*

Sketch. **Step 1:** Show that the set

$$T = \{tr\ g : g \in \Gamma, g \text{ of prime order.}\}$$

is finite.

Step 2: For any $t \in \mathbb{C}$, show that there is a normal subgroup N_t of finite index with $tr\ g \neq t$ for any $g \in N_t$.

Step 3: Let

$$N = \bigcap_{t \in T} N_t.$$

This cannot contain an element of finite order, but is finite index and normal. $\square$

Corollary 222. *Any finitely generated Kleinian group has a finite index subgroup that is torsion-free.*

Proof. Suppose Γ is Kleinian. Consider the two-to-one map $\mathrm{SL}_2(\mathbb{C}) \to \mathrm{PSL}_2(\mathbb{C})$. Let $\overline{\Gamma}$ be the preimage. By Selberg's lemma, $\overline{\Gamma}$ has a finite-index normal subgroup $\overline{N}$ that is torsion-free. We have a diagram with rising inclusions:

$$\begin{tikzcd}
\mathrm{SL}_2(\mathbb{C}) \arrow[r] & \mathrm{PSL}_2(\mathbb{C}) \\
\overline{\Gamma} \arrow[u] \arrow[r] & \Gamma \arrow[u] \\
\overline{N} \arrow[u] \arrow[r] & N \arrow[u]
\end{tikzcd}$$

Taking the two-to-one quotient preserves normality and finite index, so N is normal and finite index in Γ. It also preserves torsion-freeness. $\square$

Therefore the proof is completed by the following lemmas.

Lemma 223. *Let $\Gamma < \mathrm{SL}_2(\mathbb{C})$ be finitely generated. There are only finitely many primes p such that Γ has an element of order p.*

Proof. Suppose $\Gamma = \langle g_1, \ldots, g_m \rangle$. Suppose first that the entries of the g_i are all algebraic. Then there is a finite extension K of $\mathbb{Q}$ over which Γ is defined, say of degree n. Suppose g is of finite order p. Then the eigenvalues of g must be p-th roots of unity, hence live in $\mathbb{Q}(\zeta_p)$, an extension of degree $p-1$ over $\mathbb{Q}$. On the other hand, the eigenvalues are roots of the characteristic polynomial, so of degree at most $2n$ as algebraic numbers. Therefore $p - 1 < 2n$.

Now, suppose instead that some entries of the generators for Γ are transcendental. We will reduce to the previous case. In particular, suppose the entries include transcendentals $t_1, \ldots, t_k$, and otherwise entries live in a number field K. Then there is an ideal $\mathfrak{a}$ of $\overline{\mathbb{Q}}[X_1, \ldots, X_k]$ of polynomials vanishing on $(t_1, \ldots, t_k)$ (this actually exists: e.g. $X_1 - X_2$ vanishes on (π, π)). Choose a maximal ideal $\mathfrak{m}$ containing $\mathfrak{a}$. It must have the form

$$\mathfrak{m} = (X_1 - s_1, \ldots, X_k - s_k)$$

for some $s_i \in \overline{\mathbb{Q}}$.

Now, replace Γ with Γ' obtained by replacing t_i with s_i wherever it appears in the generating matrices. This induces a homomorphism

$$\phi : \Gamma \to \Gamma'.$$

We need only verify that whenever Γ has a point g of order p, then Γ' does also, and then apply the previous part. To verify this, note that a matrix of $\mathrm{SL}_2(\mathbb{C})$ satisfies $M^p = 1$ if and only if it satisfies $tr\ M = \zeta_p^s + \zeta_p^{-s}$ for some s (the two exponents are equal since $\det M = 1$). But as this is algebraic, it is preserved by ϕ. $\square$

Lemma 224. *Suppose* $\Gamma < \mathrm{SL}_2(\mathbb{C})$ *is finitely generated. Let* $t \in \mathbb{C}$*. Then there exists a normal subgroup* $N \lhd \mathrm{SL}_2(\mathbb{C})$ *of finite index whose elements do not have trace* t*.*

Proof. Suppose $\Gamma = \langle g_1, \ldots, g_m \rangle$. Let $R \subset \mathbb{C}$ be the subring generated by the entries of all the g_i. If $t \notin R$, we may choose $N = \Gamma$.

If $t \in R$, let $\mathfrak{m}$ be a maximal ideal not containing $t - 2$. Let K be the quotient field, which is necessarily a finite field. This induces a homomorphism to a finite group

$$\mathrm{SL}_2(R) \to \mathrm{SL}_2(K).$$

The N we seek will be the kernel of this map, restricted to Γ; this is normal and of finite index.

Now, suppose there is an element $g \in N$ with $tr\ g = 2$. Then write $g = I + h$ where h reduces to the zero matrix modulo $\mathfrak{m}$. Then $tr\ h = t - 2$ must be in $\mathfrak{m}$, a contradiction. $\square$

Selberg's lemma, among other things, motivates the definition of commensurability.

Definition 225. *Let* Γ_1 *and* Γ_2 *be Kleinian groups. Then they are* directly commensurable *if they share a finite index intersection. If* Γ_1 *is conjugate to a group commensurable to* Γ_2*, then they are* widely commensurable*.*

On the side of 3-orbifolds, wide commensurability corresponds to having a common finite hyperbolic cover. This is the more useful notion.

7.6 Fuchsian and Schottky groups

A *Schottky group* is any group obtained in the following manner. Choose disjoint circles $C_1, D_1, C_2, D_2, \ldots, C_g D_g$ in $\widehat{\mathbb{C}}$ having disjoint interiors (define the exterior as the region containing ∞; conjugate to avoid ∞ lying on a circle). Let g_i be a Möbius transformation taking C_i to D_i so that the interior of C_i is mapped to the exterior of D_i. Then $\langle g_i \rangle_{i=1}^{g}$ is a Schottky group.

Properties of any Schottky group S on g generators:

1. $S \cong F_g$, the free group on g generators

2. totally disconnected limit set (obtained by nested circle images)

3. the quotient $\Omega(S)/S$ of the ordinary set is a surface of genus g; a fundamental region is the common exterior to all circles.

4. the quotient H^3/S is a *handlebody of genus* g, i.e. a solid bagel with g holes; a fundamental region is the exterior to the geodesic planes over the circles.

Draw some pictures!

By the term *Fuchsian group*, in the context of Kleinian groups, will be meant any Kleinian group stabilizing a geodesic plane in H^3 (equivalently, a circle in $\widehat{\mathbb{C}}$). Such a group is conjugate, in $\mathrm{PSL}_2(\mathbb{C})$, to a subgroup of $\mathrm{PSL}_2(\mathbb{R})$, hence the terminology. It is of the *first kind* if its limit set is the circle itself, and of the *second kind* otherwise.

Properties of non-elementary Fuchsian groups Γ:

1. The quotient of the interior of the circle by Γ is a hyperbolic 2-manifold S; the quotient of H^3/Γ is an 'interval bundle' or cylinder over this 2-manifold, i.e. $S\times(0,1)$. To see this, imagine the case of $\mathbb{R}$: think of a plane sweeping, with axis $\mathbb{R}$, from upper half plane to lower half plane; these are the slices.

2. The boundary of the manifold differs in the first/second kind.

3. A limit set in the case of the second kind is a perfect, nowhere dense set.

An example of a Fuchsian group of the second kind is a Schottky group whose circles are all orthogonal to the real line, and whose Möbius transformations are in $\mathrm{PSL}_2(\mathbb{R})$.

Exercise 226. *Work out the limit set for the Schottky group just mentioned, explicitly (it is a Cantor-like set).*

Exercise 227. *There are exercises in Katok's book on Fuchsian groups, and also 'explorations' in Marden's book 'Outer Circles' that are appropriate. Another source is Francis Bonahon's book 'Low-Dimensional Geometry: From Euclidean Surfaces to Hyperbolic Knots'.*

7.7 Figure-Eight Knot Complement

This section will follow Chapter 11 of Francis Bonahon, *Low-Dimensional Geometry: From Euclidean Surfaces to Hyperbolic Knots.* There are a lot of pictures involved, so it is necessary to see my lecture for this material.

The relevant group will be

$$\Gamma = \left\langle \phi_1 = \begin{pmatrix} 1 & 1 \\ 1 & 1-w \end{pmatrix}, \phi_3 = \begin{pmatrix} 1 & -1 \\ -1 & 1-w \end{pmatrix}, \tau = \begin{pmatrix} 1 & w \\ 0 & 1 \end{pmatrix} \right\rangle$$

where w is a primitive cube root of unity for the remainder of this section. The fundamental domain of Γ will be described as the union

$$\Delta = \Delta_1 \cup \Delta_2$$

of two ideal tetrahedra, namely Δ_1 having vertices 0, 1, ∞, and w, and Δ_2 having vertices 0, 1, ∞ and $1 - w = w^{-1}$. These share a face $T = \Delta_1 \cap \Delta_2$ which is the ideal triangle with vertices 0, 1 and ∞. The fundamental region has 5 vertices, 9 edges, and 6 faces.

We will not prove that this is the fundamental region. However, in class we will give a gluing diagram which glues the outer faces of Δ_1 and Δ_2 to each other using the elements ϕ_1, ϕ_3 and $\tau \circ \phi_3$.

Now, let K be the figure eight knot, and let $X = \mathbb{R}^3 \setminus K$. Our goal is to prove

Theorem 228. *$X = \mathbb{R}^3 \setminus K$ is homeomorphic to H^3/Γ.*

The way to do this is to create two polyhedra, X^+ and X^-, and glue them, to create X. We will then compare the gluing data to the fundamental domain gluing data to see that they agree.

7.8 Bianchi groups and fundamental domains

Let K be an imaginary quadratic field with ring of integers $\mathcal{O}_K$. Then $\mathrm{PSL}_2(\mathcal{O}_K)$ is a Kleinian group (we showed this long ago). We'd like to consider their fundamental domains now.

First, note that translations by the elements of the ring of integers form a subgroup of parabolics that fixes ∞ – a lattice of rank 2 in fact – so ∞ is a cusp and these are not co-compact.

Theorem 229. *The orbits of $\widehat{K} = K \cup \{\infty\}$ are in bijection with the ideal classes of $\mathcal{O}_K$. Furthermore, the number of Γ-equivalence classes of cusps is equal to the class number of $\mathcal{O}_K$.*

Let us quickly recall the definition of ideal classes. Let $\mathcal{I}(\mathcal{O}_K)$ denote the non-zero ideals of $\mathcal{O}_K$. Two ideals $\mathfrak{a}$ and $\mathfrak{b}$ of $\mathcal{O}_K$ are equivalent if $\mathfrak{a} \cdot a\mathcal{O}_K = \mathfrak{b} \cdot b\mathcal{O}_K$ for some non-zero $a, b \in \mathcal{O}_K$. The *ideal class group* of $\mathcal{O}_K$ is the collection of non-zero ideal classes under ideal multiplication: this is a group. We denote it $\mathcal{P}(\mathcal{O}_K)$.

Proof. Write $\Gamma = \mathrm{PSL}_2(\mathcal{O}_K)$.

First, we show that the cusps of Γ occur exactly at the elements of $\widehat{K}$.

Let $k \in K$. There is some element $h \in \mathrm{PSL}_2(K)$ such that $h(\infty) = k$. We can write $h = \frac{1}{d} f$ for some $f \in \mathrm{PGL}_2(\mathcal{O}_K)$ and some non-zero $d \in \mathcal{O}_K$. Therefore, we have

$$hgh^{-1} \in \mathrm{PSL}_2(\mathcal{O}_K) \text{ whenever } g \in \mathrm{PSL}_2(\mathcal{O}_K).$$

Therefore, the stabilizer of k is isomorphic to the stabilizer of ∞ and k is a cusp.

Now, suppose $k \in \widehat{\mathbb{C}}$ is a cusp. Then there is a parabolic element of Γ which fixes k. But then k is a repeated root of a quadratic equation over K (i.e. giving the fixed points of the parabolic), which implies that $k \in K$.

Next, we show that cusps are Γ-equivalent if the corresponding elements of $\mathbb{P}^1(\mathcal{O}_K)$ represent the same ideal class, in the sense of the well-defined map

$$\mathbb{P}^1(\mathcal{O}_K) \to \mathcal{P}(\mathcal{O}_K), \quad [\alpha, \beta] \mapsto (\alpha, \beta)\mathcal{O}_K.$$

The points $[\alpha, \beta]$ and $[\gamma, \delta]$ are in the same orbit if and only if

$$s\alpha = t(a\gamma + b\delta), s\beta = t(c\gamma + d\delta)$$

where $a, b, c, d \in \mathcal{O}_K,\ ad - bc = 1,\ s, t \in \mathcal{O}_K,\ st \neq 0$. But this happens if and only if we have the homothety of lattices

$$s(\alpha\mathcal{O}_K + \beta\mathcal{O}_K) = t(\gamma\mathcal{O}_K + \delta\mathcal{O}_K).$$

or in other words, equality of ideal classes

$$[(\alpha, \beta)\mathcal{O}_K] = [(\gamma, \delta)\mathcal{O}_K].$$

Therefore the orbits of $\widehat{K}$ are in bijection with the ideal classes of $\mathcal{O}_K$. $\square$

In order to find a fundamental domain, we will define the notion of a *Ford fundamental domain*.

First, the *isometric circle* I_g of g is that circle in $\widehat{\mathbb{C}}$ upon which $|g'(z)| = 1$. In particular, since $g'(z) = (cz + d)^{-2}$, it is exactly the circle

$$|z + d/c| = 1/|c|, \text{ or } |cz + d|^2 = 1.$$

The *isometric sphere* or *isometric plane* is the hemisphere lying above this circle, i.e.

$$|z + tj + d/c| = 1/|c|, \text{ or } |cz + d|^2 + |c|^2t^2 = 1.$$

Warning: the isometric circle and plane depend on the model we are using: the upper half space model. They are not intrinsinc, in the sense that an isometry between models (or even back to the same model) may not conjugate the isometric circles as it would the axes and fixed points.

One of the important geometric properties of isometric circles is that g interchanges I_g and $I_{g^{-1}}$, swapping interiors with exteriors. For, the image of I_g is $|g'(g^{-1}(z))| = 1$. But $I_{g^{-1}}$ is $|(g^{-1})'(z)| = 1$. These are equal.

We have

1. $tr\ ^2 g = 0$ (elliptic order two) if and only if $I_g = I_{g^{-1}}$

2. $0 < |tr\ ^2 g| < 4$ if and only if I_g and $I_{g^{-1}}$ intersect at two points (the axis of rotation if g is elliptic).

3. parabolic if and only if I_g and $I_{g^{-1}}$ are tangent

4. $|tr\ ^2 g| > 4$ if and only if I_g and $_{g^{-1}}$ are disjoint.

The largest isometric circles for $\mathrm{PSL}_2(\mathcal{O}_K)$ form a repeating pattern, one of radius 1 above each algebraic integer. In the case $\mathcal{O}_K = \mathbb{Z}[i]$, all other isometric circles lie below these.

The stabilizer at ∞ has a fundamental region tesselation given by planes above $\widehat{\mathbb{R}} + a + 1/2$ and $i\widehat{\mathbb{R}} + ia/2$ for $a \in \mathbb{Z}$. The reason this is not just a square is that the stabilizer at ∞ also includes the map $z \mapsto -z$, a rotation by π. Intersecting this with the area exterior to all isometric circles gives what is called a *Ford domain.* In the case of $\mathbb{Z}[i]$ it is, explicitly,

$$\{z + tj : |z|^2 + t^2 > 1, -1/2 < \Re(z) < 1/2, -1/2 < \Im(z) < 0\}.$$

A general theorem on Ford domains shows that this is a fundamental domain for $\mathrm{PSL}_2(\mathbb{Z}[i])$. Alternatively, or, if we wish to turn this into a presentation for the group, we can use Poincaré's Polyhedron Theorem, which I will now describe.

7.9 Poincaré's Polyhedron Theorem

One source of Kleinian groups is to construct tessellations of H^3 and determine the associated group. There is a set of conditions under which a given polyhedron will tessellate.

Recall that a polygon is a region in a geodesic plane whose boundary is made up of finitely many geodesic segments, meeting in pairs at their endpoints. A polyhedron is a region with a boundary made up of faces which are polygons in geodesic planes. Each edge is adjacent to exactly two faces. These are required to include their boundaries, so they are closed in $H^3 \cup \widehat{\mathbb{C}}$.

Vertices lying on $\widehat{\mathbb{C}}$ are allowed and called *ideal.* It is also possible that faces touch $\widehat{\mathbb{C}}$ along circle arcs.

By *gluing data* we will mean that the faces of the polyhedron are paired

$$\{F_1, F_1'\}, \{F_2, F_2'\}, \ldots$$

along with isometries

$$\phi_i : F_i \to F_i', \quad \phi_i^{-1} : F_i' \to F_i.$$

Our goal is to understand, first, when a given polyhedron may tile H^3, i.e. form a *tessellation.*

Definition 230. *A* tessellation *of H^3 is given by the collection of images of a polyhedron X with gluing data whenever*

1. *the images of X under the* tiling group *generated by the ϕ_i cover H^3*
2. *the intersection of any two images consists of shared complete vertices, edges and faces (i.e. not parts of edges or faces)*
3. *around any point $\alpha \in H^3$, there is an open ball meeting only finitely many of the images.*

The last part is *local finiteness*, which we've encountered before.

The following is immediate:

Proposition 231. *If images do not repeat, then X forms a fundamental domain for the associated tiling group.*

We can now state a general theorem about when a polyhedron will tessellate.

Theorem 232. *Let X be a connected polyhedron in H^3, with gluing data as above. Suppose that the following conditions hold.*

1. *(Dihedral angle condition) Let E be an edge of X. Then*
$$\sum_{F \sim E} \textit{dihedral angle of } X \textit{ along } F = 2\pi/n$$
where $F \sim E$ indicates that edges F and E of X are glued, and n is some integer.
2. *(Edge orientation condition) The edges of X can be oriented so that gluing preserves orientation.*
3. *(Horosphere condition) For every ideal vertex, we can associated a small horosphere at that vertex in such a way that gluing ideal vertices maps the horospheres to one another.*

Then, the polyhedron tessellates H^3. Furthermore, the associated tiling group is discrete.

Edge orientation, if it fails, can generally be repaired by adding extra vertices. Orient infinite edges toward the boundary. For all other edges, split them by a vertex at the midpoint and orient outward. (Here, midpoint can be determined by using the horospheres to cut off the infinite ends, making it finite.)

Now we can show that the fundamental domain for $\mathrm{PSL}_2(\mathbb{Z}[i])$ does indeed tessellate. This requires diagrams and I'm doing it in class, but not here in the notes.

Now, the side-pairings on this fundamental domain, as it turns out, do indeed give us a presentation of $\mathrm{PSL}_2(\mathcal{O}_K)$:

$\mathrm{PSL}_2(\mathbb{Z}[i]) = \langle x, y, z, w : x^2 = y^2 = w^2 = 1, (zx)^3 = (zy)^2 = (zw)^2 = 1, (yx)^2 = 1, (wx)^3 = 1 \rangle.$

In fact, we can take as generators
$$x = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, y = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}, z = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, w = \begin{pmatrix} i & -1 \\ 0 & -i \end{pmatrix}.$$
These identify various sides of the fundamental domain.

8 Review of Algebraic Number Theory

Topics:

1. Number fields: definition, primitive element theorem, minimal polynomials, real and complex places, Galois vs. not, conjugates of an element, norm and trace, discriminant
2. Algebraic numbers/integers, equivalent conditions for being integrally closed, integral bases and discriminant,
3. Ideals in the ring of integers, fractional ideals, ideal class group, norm of ideals, splitting of primes, residue field
4. units, Dirichlet's unit theorem, class groups (overview of what's known),
5. quadratic fields as an example of all the above, non-maximal orders in quadratic fields and conductor

9 Quaternion algebras

This follows Section 2 of Maclachlan and Reid. Since we are following decently closely, I recommend all the exercises in this section of the book.

9.1 Important Warning

Throughout this section we will assume K is a field of characteristic $\neq 2$. This will be important in the definition of a quaternion algebra and elsewhere; see Maclachlan and Reid.

9.2 Algebras

A particularly concrete way to define algebras is the following. We will always use associative unital finite-dimensional algebras.

Definition 233. *An* algebra B *over a field* K *(or* K-algebra*) is a finite dimensional vector space equipped with an associative bilinear map* $B \times B \to B$ *with unity.*

One can check that the set of axioms implied by this definition make B into an associative unital ring containing K as a subring in its centre (the centre is the subring of elements that commute with everything under multiplication). Therefore one sometimes makes the following equivalent definition.

Definition 234. *An* algebra B *over a field* K *(or* K-algebra*) is an associative unital ring, together with an embedding* $f : K \to B$ *taking* K *into the centre of* B*, and such that* B *has finite dimension over* K*.*

Exercise 235. *Show that these are equivalent definitions.*

A particularly basic example of an algebra is any field extension of K of finite degree, and a matrix algebra $M_n(K)$, the $n \times n$ matrices over K (which has dimension n^2).

Definition 236. *A homomorphism of* K*-algebras is a ring homomorphism that restricts to the identity on* K*. The* K*-algebra of* K*-algebra homomorphisms* $B \to B$ *is denoted* $\mathrm{End}_K(B)$ *(it is a ring under pointwise addition and composition) and its invertible elements (isomorphisms) form a group denoted* $\mathrm{Aut}_K(B)$*.*

As usual, check the various assertions made in the definition about the algebraic structures on these new objects. Note that since a K-algebra is in particular a vector space, its endomorphisms are in particular linear transformations. We may also refer to the *dimension* of a K-algebra when needed.

9.3 The Hamilton Quaternions

The *Hamilton quaternions* over a field K, denoted $\mathcal{H}$, are the four-dimensional algebra over K with basis $1, i, j, k$ where the bilinear product is determined by the relations $i^2 = j^2 = -1$ and $k = ij = -ji$.

This ring is not commutative.

Exercise 237. *How is* $\mathcal{H}$ *obtained from the quaternion group (*Q_8 *in Dummit and Foote)?*

When $K = \mathbb{C}$, $\mathcal{H}$ has a standard representation $\mathcal{H} \to M_2(\mathbb{C})$, given by

$$1 \mapsto \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad i \mapsto \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}, \quad j \mapsto \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad k \mapsto \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}.$$

(We are abusing notation by using two different i's.) This is not a representation in the sense of group theory, but of algebras: a pair (V, ρ) where V is a vector space over K, and ρ is a map from the algebra to the endomorphisms of V. In the example above, we have $V = \mathbb{C}^2$ and ρ given by its values on $1, i, j, k$.

Exercise 238. *We could loosen the definition of algebra to consider infinite-dimensional vectors spaces. Show that* $\mathbb{C}[x]$*, where* x *is an indeterminate, is such an algebra. What are its finite-dimensional representations?*

Exercise 239. *Explain the relationship between representations of group algebras and representations of groups.*

9.4 Hilbert symbol

Generalizing the Hamilton quaternions, for $a, b \in K^*$, we define $\left(\frac{a,b}{K}\right)$ (called the *Hilbert symbol*) to be the four-dimensional algebra over K with basis $1, i, j, k$ where the bilinear product is determined by the relations $i^2 = a$, $j^2 = b$ and $k = ij = -ji$.

For example, the Hamilton quaternions are $\left(\frac{-1,-1}{K}\right)$.

However, different Hilbert symbols may give isomorphic algebras:

Exercise 240. $\left(\frac{a,b}{K}\right) \cong \left(\frac{b,a}{K}\right) \cong \left(\frac{a,-ab}{K}\right) \cong \left(\frac{b,-ab}{K}\right)$. *Also* $\left(\frac{1,-1}{K}\right) \cong \left(\frac{1,1}{K}\right) \cong M_2(K)$. *Give the isomorphisms explicitly.*

In particular, there are various different choices for the basis of any quaternion algebra, but we will always assume the first basis vector is $1 \in K$.

quat-alg-closed **Proposition 241.** *Let $a, b \in K^*$. If K is algebraically closed, then*

$$\left(\frac{a,b}{K}\right) \cong M_2(K).$$

Proof. First, note that for any field K,

$$\left(\frac{1,1}{K}\right) \cong M_2(K)$$

by the representation

$$1 \mapsto \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad i \mapsto \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad j \mapsto \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad k \mapsto \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

Therefore it suffices to check that

$$\left(\frac{x^2, y^2}{K}\right) \cong \left(\frac{1,1}{K}\right)$$

for any $x, y \in K^*$. Since K is algebraically closed, the result will follow from the observation that every element is a square. But we can use the isomorphism

$$\phi : \left(\frac{1,1}{K}\right) \to \left(\frac{x^2, y^2}{K}\right)$$

given by

$$\phi(1) = 1, \quad \phi(i) = xi, \quad \phi(j) = yj, \quad \phi(k) = xyk.$$

$\square$

The definition is functorial in K, in the sense that $\left(\frac{a,b}{K}\right) \otimes_K F \cong \left(\frac{a,b}{F}\right)$ whenever F is a field containing K.

Any quaternion algebra A, as a vector space, has a decomposition $A \cong K \oplus A_0$, where A_0 is called the *pure quaternions*, and consists of 0 together with the elements not in K whose squares are in K (this can be verified by a computation). In practice, it is the span of $\{i, j, k\}$, but it does not depend on the choice of quaternion basis.

Definition 242. *Let A be a quaternion algebra. There is a map $\bar{\cdot} : A \to A$ called* conjugation *given by $\overline{a + \alpha} = a - \alpha$ whenever $a \in K$ and $\alpha \in A_0$.*

It has the following properties, for any $x, y \in A$ and $k \in K$ (check them):

1. $\overline{kx} = k\overline{x}$ and $\overline{x + y} = \overline{x} + \overline{y}$ (linear transformation)
2. $\overline{xy} = \overline{y}\,\overline{x}$ (anti-automorphism)
3. $\overline{\overline{x}} = x$ (involution)
4. $x\overline{x} \in K$ ('standard')

In light of the above it is called a *standard anti-involution*.

Definition 243. *Let A be a quaternion algebra. The* norm *$n : A \to K$ is the map $n(x) = x\overline{x}$. The* trace *$tr : A \to K$ is the map $tr(x) = x + \overline{x}$.*

For $A = \left(\frac{a,b}{K}\right)$, we have

$$n(a_0 + a_1 i + a_2 j + a_3 k) = a_0^2 - aa_1^2 - ba_2^2 + aba_3^2.$$

The norm is multiplicative, and the trace is additive (check it). Also check that $tr(xy) = tr(yx)$ and that $x^2 - tr(x)x + n(x) = 0$.

Write A^* for the invertible elements of A and A^1 for the elements of norm 1. Since if $n(x) \neq 0$, we have $x\frac{\overline{x}}{n(x)} = 1$, we have

$$A^* = \{x \in A : n(x) \neq 0\}.$$

Exercise 244. *From a previous exercise, $\left(\frac{1,1}{K}\right) \cong M_2(K)$. Determine the conjugation, norm and trace on the matrix algebra $M_2(K)$.*

Exercise 245. *If B is any K-algebra, then there is an injection $B \to \mathrm{End_K}(B)$, i.e. $x \in B$ can be identified with its left multiplication action on B. Endomorphisms have trace and determinant; how are these related to the trace and determinant as defined above in the case that B is a quaternion algebra?*

9.5 Centrality and Simplicity

Definition 246. *A K-algebra B is* simple *if it has no proper two-sided ideals. It is* central *if its centre, $Z(B)$, is exactly equal to K.*

Note that two-sided ideals of B must also be subspaces. We will just say *ideals* for *two-sided ideals.*

Exercise 247. *Give an example of a K-algebra with a subspace which is not a two-sided ideal.*

Exercise 248. *Show that $M_n(K)$ is central and simple.* Hint: To show simplicity, assume an ideal contains a matrix with some non-zero entry. Use absorption to show that this is enough to cause the ideal to contain the entire ring (this relies on the fact that K is a field). To show centrality, first make sure you know how K embeds in $M_n(K)$, then try commuting with elementary matrices. *Give an example, however, that $M_n(K)$ does have proper one-sided ideals.*

Proposition 249. *Quaternion algebras are central and simple.*

Proof. Let $A = \left(\frac{a,b}{K}\right)$. The key to the proof is to extend scalars to the algebraic closure $\overline{K}$ of K; this idea will come up again soon.

For, by a previous exercise, $A \otimes_K \overline{K} \cong M_2(\overline{K})$, which we know is central. Therefore, restricting scalars, we obtain that A is central.

Now, as for simplicity. Any two-sided ideal of A will extend to a two-sided ideal of $A \otimes_K \overline{K} \cong M_2(\overline{K})$ preserving its dimension; but the latter is simple, so the extended ideal is of dimension 0 or 4. Hence A is not proper. $\square$

Exercise 250. *The non-zero pure quaternions are exactly the non-central elements whose squares are central. This is Lemma 2.1.4 in Maclachlan and Reid.*

9.6 Brauer Groups, Wedderburn and Skolem Noether – in brief

Proposition 251. *If A and B are central simple, then so $A \otimes B$.*

Exercise 252. *Provide a proof. This is an excellent exercises for making sure you know what the tensor product is. It is Proposition 2.8.4 in Maclachlan and Reid.*

Exercise 253. *An element of a K-algebra B can be considered an endomorphism on itself by left multiplication; this is called the* left regular representation.

1. *One must be careful though, this phrase might mean $B \to \mathrm{End}_B(B)$, where the latter is the right B-module of B-module endomorphisms of B; show that this is an isomorphism.*

2. *Or it could mean $B \to \mathrm{End}_K(B)$. Show that this map is an injection but not an isomorphism.*

3. *One also has the related fact that $B^o \cong \mathrm{End}_B(B)$, if you take the latter to be a* left *B-module. (Here B^o is the opposite ring, i.e. replace multiplication $a \cdot b = ab$ with the alternate version $a \cdot b = ba$.) Show this.*

4. *Finally, show that* if B is a central simple K-algebra*, then $B \otimes B^o \cong \mathrm{End}_K(B)$.*

The collection of central simple K-algebras, under a certain equivalence, forms a group, $\mathbf{Br}(K)$, the *Brauer group of K*. The equivalence relation is given by

$$A \sim B \iff A \otimes M_n(K) \cong B \otimes M_m(K)$$

for some m, n. The operation is given by

$$[A][B] = [A \otimes B].$$

Notice the similarity to the ideal class group of a field. Here tensor product is a natural product on central simple K-algebras, and matrix algebras play the role of principal ideals. The inverse of an algebra A is A^o, its opposite algebra, as in the last exercise.

Exercise 254. *Check everything I just claimed: i.e. the equivalence relation is an equivalence relation, the group operation is a group operation, inverse is what I claim, etc.*

Now, we will briefly survey the two big theorems: the Skolem Noether Theorem and Wedderburn's Structure Theorem.

Theorem 255 (Wedderburn's Structure Theorem)**.** *Let A be a simple K-algebra. Then $A \cong M_n(D)$ for some integer n and $D \cong \mathrm{End}_A(N)$, a division algebra, where N is a minimal right ideal N of A. These n and D are unique.*

Theorem 256 (Skolem Noether Theorem)**.** *Let A be a central simple K-algebra, and let B be a simple K-algebra. Let $\phi, \psi : B \to A$ be algebra homomorphisms. Then there is some $c \in A^*$ so that $\phi = c^{-1}\psi c$. In particular, every non-zero endomorphism of A is inner.*

9.7 Characterising Quaternion Algebras

Definition 257. *A* division algebra *is an algebra in which every non-zero element is invertible.*

The reason a division algebra needn't be a field is the possibility of non-commutativity. Evidently $M_n(K)$ is not a division algebra. We previously verified that the Hamilton quaternions over $\mathbb{R}$ form a division algebra.

Exercise 258. *Show that a commutative (associative unital finite-dimensional) algebra is a field.*

We have seen that quaternion algebras are central simple K-algebras of dimension 4. We now have the converse, and see that they are all either division algebras or matrix algebras.

Theorem 259. *Every central simple K-algebra of dimension 4 is a quaternion algebra. These come in two varieties: division algebras and $M_2(K)$.*

Proof. This proof is a rephrasing of that from Maclachlan and Reid, Theorem 2.1.7-8. First we show that a simple 4-dimensional algebra over a field K is either a division algebra or isomorphic to $M_2(K)$. This is a consequence of Wedderburn's Structure Theorem: a simple algebra is isomorphic to $M_n(D)$ for some division algebra D. Therefore $n^2m = 4$ where D has dimension m. If $m = 1$, we are in the case of $M_2(K)$. The only other possible dimension is $m = 4$, when we are in the case $M_1(D) \cong D$.

Let A be a central simple K-algebra of dimension 4. If $A \cong M_2(K)$, it is a quaternion algebra. So we may assume instead that it is a division algebra. The proof proceeds by finding a basis of elements $1, y, z, yz$ such that $y^2, z^2 \in K$.

Let $w \in A \setminus K$. Then, as observed above, w satisfies a quadratic equation (in terms of norm and trace). Now $K(w)$ (the K-algebra generated by $1, w$) is therefore a quadratic field extension of K. Therefore there is some element $y \in A$ such that $y^2 \in K$. Write $a := y^2$.

Now conjugation $y \mapsto -y$ on $K(w)$ is induced by conjugation by some non-zero element z of A, by the Skolem Noether Theorem, i.e. $zyz^{-1} = -y$.

Next, we claim that $1, y, z, yz$ forms a basis for A. This follows from the observation that since $z \notin K(w)$, the $K(w)$-algebra spanned by $1, z$ must be of dimension ≥ 2 over $K(w)$, hence of dimension ≥ 4 over K.

Next, $z^2yz^{-2} = y$ so that z^2 is in the centre of A, hence $b := z^2 \in K$ (this uses centrality of A).

Therefore, $A \cong \left(\frac{a,b}{K}\right)$. □

Recall that we are assuming $\mathrm{char}(K) \neq 2$. See Maclachlan and Reid for comments.

9.8 Orders in Quaternion Algebras

Important: As in Maclachlan and Reid, Section 2.2, we will let R be a Dedekind domain (integrally closed Noetherian integral domain in which non-zero primes are maximal) with fraction field k which is either a number field or p-adic (or $\mathfrak{p}$-adic) field. We will let A be a quaternion algebra over k.

Examples:

1. k a number field, R its ring of integers

2. $k = \mathbb{Q}_p$, $R = \mathbb{Z}_p$, i.e. the p-adic case

3. $k = K_\mathfrak{p}$, $R = (\mathcal{O}_K)_\mathfrak{p}$, for $\mathfrak{p}$ a prime ideal of a number field K, i.e. the $\mathfrak{p}$-adic case.

4. $k = \mathbb{Q}$, $R = \mathbb{Z}_{(p)}$, the localization of $\mathbb{Z}$ at the prime ideal (p), i.e. no p's in denominators

You can mainly keep in mind the first two. In (2), (3), (4), R is a discrete valuation ring, which is a simple type of Dedekind domain.

Definition 260. *Let V be a vector space over k. A subset L is an R-*lattice *if it is a finitely generated R-module. An R-lattice L is* complete *if $L \otimes_R k = V$.*

As k has characteristic zero by assumption, such lattices are torsion-free.
Examples:

1. $R^n \subseteq k^n$ (e.g. $\mathbb{Z}^2 \subseteq \mathbb{Q}^2$)

2. ring of integers in a number field: $\mathcal{O}_N \subseteq N$ (e.g. $\mathbb{Z}[i] \subseteq \mathbb{Q}(i)$)

3. $M_2(R) \subseteq M_2(k) \cong k^4$

4. $\mathbb{Z}[i, j]$ (integer Hamilton quaternions) inside $\left(\frac{-1,-1}{\mathbb{Q}}\right)$ (rational Hamilton quaternions)

Recall that if N is a number field (which is a vector space over $\mathbb{Q}$), then $\alpha \in \mathcal{O}_N$ is an *integer* if $\mathbb{Z}[\alpha]$ is a complete $\mathbb{Z}$-lattice in N. In greater generality, we have the following.

Definition 261. *Let A be a quaternion algebra or a number field over k. Then $\alpha \in A$ is an* integer *(i.e.* integral over *R) if $R[\alpha]$ is a R-lattice in A.*

Proposition 262. *Let A be a quaternion algebra. Then $\alpha \in A$ is an integer if and only if $tr\ (\alpha), n(\alpha) \in R$.*

Proof. (Maclachlan and Reid, Lemma 2.2.4, rephrased) If $\alpha \in A$, then α satisfies

$$\alpha^2 - tr\ (\alpha) + n(\alpha) = 0.$$

In particular, if $tr\ (\alpha), n(\alpha) \in R$, then $\alpha^2 \in R + R\alpha$, which implies that $R[\alpha] = R + R\alpha$. Therefore, $R[\alpha]$ is a finitely generated R-module.

Conversely, suppose $\alpha \in A$ is an integer, i.e. $R[\alpha]$ is a finitely generated R-module. We must consider two cases.

First, suppose A is a division algebra. Then $N = k(\alpha)$ is a quadratic extension (unless $\alpha \in R$, in which case we are done), and α is in the integral closure of R in N. The quaternion algebra trace and norm becomes the field trace and norm, and we are done.

Second, suppose $A \cong M_2(k)$. Then α can be conjugated to be of the form

$$\alpha = \begin{pmatrix} a & b \\ 0 & c \end{pmatrix}, \quad \alpha^n = \begin{pmatrix} a^n & d_n \\ 0 & c^n \end{pmatrix}, \quad a, b, c, d_n \in k.$$

In particular, then, if $R[\alpha]$ is a finitely generated R-module, then restricting to the first matrix entry gives a finitely generated R-module. Therefore $a \in k$ is integral over R, hence $a \in R$. Similarly, $c \in R$. This implies $tr\ (\alpha), n(\alpha) \in R$. $\square$

The last theorem is actually a special case of this more general theorem in commutative algebra, the proof of which is worth knowing.

thm: intequiv

Theorem 263. *Let A be a ring with subring R and suppose $\alpha \in A$. The following are equivalent:*

1. *α satisfies a monic irreducible polynomial with coefficients in R.*
2. *$R[\alpha]$ is a finitely generated R-module.*
3. *There exists a subring $B \subset A$ such that*
   (a) *$R \subset B$*
   (b) *$\alpha \in B$*
   (c) *B is a finitely generated R-module*

Proof. 1 $\implies$ 2: Let α satisfy a monic irreducible f of degree n over R. Let $M = R + R\alpha + \cdots + R\alpha^{n-1}$. Then $\alpha^n = -f(\alpha) + \alpha^n \in M$ and similarly $\alpha^{n+j} \in M$ for all $j \geq 0$. So $M = R[\alpha]$ is finitely generated as an R-module.

2 $\implies$ 3: Immediate.

3 $\implies$ 1: $B = Ry_1 + \cdots + Ry_n$ for some y_i. Then $\alpha y_i \in B$ so it has the form

$$\alpha y_i = \sum_{j=1}^{n} a_{ij} y_j$$

i.e., for each i,

$$\sum_{j=1}^{n} (\alpha\delta_{ij} - a_{ij}) y_j = 0.$$

Consider the matrix $M = (\alpha\delta_{ij} - a_{ij})_{i,j}$. Then $M^{adj} M = \det(M) I$ where *adj* denotes adjugate (the adjugate is the matrix satsifying this equation by definition). Since $M\mathbf{y} = \mathbf{0}$, then $det(M) y_i = 0$ for all y_i and so $\det(M) B = 0$. In particular, $\det(M) \cdot 1 = 0$ which implies $\det(M) = 0$. But $\det(M)$ is a monic polynomial of degree n in α, with coefficients in R. $\square$

The integers in a quaternion algebra need not form a ring!

Example from Maclachlan and Reid, constructed using a Pythagorean triple: let $A = \left(\frac{-1,3}{\mathbb{Q}}\right)$.

1. $\alpha := j$ is an integer, since $tr\ (j) = 0$ and $n(j) = 1$.
2. $\beta := \frac{3j+4ij}{5}$ is an integer, since $tr\ (\beta) = 0$ and $n(\beta) = -3 \cdot 3^2/5^2 - 3 \cdot 4^2/5^2 = -3$
3. $\alpha + \beta$ is not an integer, since $tr\ (\alpha+\beta) = 0$ but $n(\alpha+\beta) = -3 \cdot 8^2/5^2 - 3 \cdot 4^2/5^2 = -3(8^2+4^2)/5^2$ is not an integer.

Example from Pete L. Clark's notes (corrected): let $A = M_2(\mathbb{Q})$.

1. $\alpha := \begin{pmatrix} 1/2 & -3 \\ 1/4 & 1/2 \end{pmatrix}$ is an integer, since $tr\ (\alpha) = 1$, $n(\alpha) = \det(\alpha) =?$.
2. $\beta := \begin{pmatrix} 0 & 1/5 \\ 5 & 0 \end{pmatrix}$ is an integer, since $tr\ (\beta) = 0$, $n(\beta) = \det(\beta) = -1$.
3. $\alpha\beta$ is not integral, since it has trace $1/20 - 15$.
4. $\alpha + \beta$ is not integral, as its determinant is $1/4 + 147/10$.

Definition 264. *A complete R-lattice in a quaternion algebra A over k is called an* ideal *of A. It is called an* integral ideal *if it consists of integral elements.*

It is useful to compare with the number field case: let N be a number field, and $\mathcal{O}_N$ its ring of integers. Any complete $\mathbb{Z}$-lattice L of N has an order, defined as $\text{ord}(L) = \{\alpha \in N : \alpha L \subset L\}$. Note that $\text{ord}(L)$ is always a subring of $\mathcal{O}_N$. The fractional ideals of $\mathcal{O}_N$ are exactly those complete $\mathbb{Z}$-lattices L of N with $\text{ord}(L) = \mathcal{O}_N$. So the notion of an ideal of A is analogous to the notion of a fractional ideal of some order of N.

Examples of quaternion algebra ideals include $R[e_1, e_2, e_3, e_4]$ where the e_i form any k-basis of A.

Definition 265. *An ideal of A which is also a ring containing* 1 *is called an* order *of A.*

Examples of orders include the following:

1. When $A \cong \left(\frac{a,b}{k}\right)$ with basis $1, i, j, ij$, then $R[i,j] = R[1,i,j,ij]$ is an order.
2. $M_2(R)$ is an order in $M_2(k)$

Proposition 266. *Let A be a quaternion algebra. A subset $\mathcal{O} \subset A$ is an order if and only if it is a ring, consisting entirely of integers, containing R and such that $k\mathcal{O} = A$.*

The proof gives me an excuse to emphasize the importance of the trace pairing to number theory and algebra. First, note that the endomorphisms of a vector space V have the form

$$\mathrm{End}(V) \cong V^* \otimes V$$

where V^* is the dual, the space of linear functionals. Choose a standard basis e_i and write your linear transformation with respect to it. Think of the rows of the corresponding matrix as linear functionals $f_1, \ldots, f_n$. Then the corresponding linear transformation is accomplished by

$$(\sum f_i \otimes e_i)(v) = \sum f_i(v)e_i.$$

The map

$$V^* \times V \to k, \quad (f, v) \mapsto f(v)$$

is surely very natural; it is bilinear and extends to a map $\mathrm{End}(V) \cong V^* \otimes V \to k$. In fact, this is the trace map (convince yourself of this right now).

Now, given this linear map $tr\ : \mathrm{End}(V) \to k$, we construct a bilinear form

$$\langle \cdot, \cdot \rangle : \mathrm{End}(V) \times \mathrm{End}(V) \to k, \quad \langle S, T\rangle = tr\ (ST).$$

The structure of algebraic number fields and quaternion algebras is first and foremost that of a vector space. In particular, multiplication in the ring $x \mapsto \alpha x$ is a linear transformation. In fact, we have an injection

$$N \to \mathrm{End}(N).$$

This gives us the trace pairing on $N \times N$, given by $\langle x, y\rangle = tr\ (xy)$.

We will also need to discuss the notion of a dual lattice with respect to a bilinear pairing. Suppose we have a vector space V over k and a bilinear pairing $\langle \cdot, \cdot \rangle : V \times V \to k$. We already have the notion of the dual vector space V^*, but with this choice of pairing, we can identify V with V^* by $x \mapsto \langle x, \cdot \rangle$.

Let L be an R-lattice in V. Then we define

$$L^\vee = \{\mathbf{x} \in V : \langle \mathbf{x}, \mathbf{y}\rangle \in R \ \ \forall \mathbf{y} \in L\}.$$

The set $L^\vee$ consists of linear functionals $V \to R$.

Choose coordinates so we can write matrices and write G_M for the Gram matrix for the pairing. Suppose we choose a basis for L, and use it as columns in a matrix B. Then

$$\begin{aligned} &\mathbf{v} \in L^\vee \\ &\iff B^T G_M \mathbf{v} \in R^n \\ &\iff \mathbf{v} \in G_M^{-1} B^{-T} R^n \\ &\iff \mathbf{v} \in BG^{-1} R^n \end{aligned}$$

where $G := B^T G_M B$. In particular, if the lattice takes integral values under the bilinear form, then

$$L^\vee \subseteq \frac{1}{\det(G)} L.$$

To say this in a more coordinate-free way (still identifying V with V^* via the bilinear pairing), one should let $f : R^n \to L$ be an isomorphism; this has an adjoint $g : L \to R^n$ with respect to the pairing (i.e. it is defined by the behaviour $\langle f(x), y\rangle = \langle x, g(y)\rangle$). Then $L^\vee = g^{-1}(R^n)$.

Proof. Suppose $\mathcal{O}$ is an order. Then it is a ring containing R by definition. As a complete R-lattice, we have $k\mathcal{O} = A$. It remains to show that it consists of integers. Let $\alpha \in \mathcal{O}$. Then $R[\alpha]$ must be an R-lattice, since $\mathcal{O}$ is, so α is integral.

For the converse, suppose $\mathcal{O}$ is a ring, consisting entirely of integers, containing R, and such that $k\mathcal{O} = A$. All that is required to prove is that $\mathcal{O}$ is finitely generated. Since $k\mathcal{O} = A$, choose a basis for A made of elements of $\mathcal{O}$. Let L be the R-module generated by this basis. Then $\mathcal{O} \subseteq L^\vee \subseteq \frac{1}{d}L$ where d is the determinant of the trace pairing on L. But then, since L is finitely generated, so is $\frac{1}{d}L$ and hence so is $\mathcal{O}$ (since R is Noetherian). $\square$

We also have the notion of order in rings of integers: an *order* of a number field N is a complete $\mathbb{Z}$-lattice of N which is a ring containing 1. For any complete $\mathbb{Z}$-lattice L, $\mathrm{ord}(L)$ is an order. This motivates this definition:

Definition 267. *Let I be an ideal. Then it has a* left *and* right order*, respectively:*

$$\mathcal{O}_\ell(I) = \{\alpha \in A : \alpha L \subseteq L\}, \quad \mathcal{O}_r(I) = \{\alpha \in A : L\alpha \subseteq L\}.$$

Proposition 268. *Left and right orders are actually orders.*

Proof. Let $\mathcal{O}$ be a left or right order. Then $\mathcal{O}$ is certainly a ring containing R. Furthermore, $k\mathcal{O} = A$ since any $\alpha \in A$ satisfies $\alpha^{-1}\alpha I \subseteq I$. In other words, $\alpha/x \in \mathcal{O}$ for some $x \in k$.

Next we show any $\alpha \in \mathcal{O}$ is an integer. Let $x \in I$. Since R is Noetherian, $R[\alpha]x$ is finitely generated as an R-submodule of I, which is finitely generated. Therefore $R[\alpha]$ is finitely generated. $\square$

In the quaternion or number field case, orders are ordered by inclusion, and there is a maximal order (by Zorn's Lemma). In the case of the number field, the maximal order is the ring of integers. In a quaternion algebra, there are maximal orders, possibly more than one.

9.9 Orders in $M_2(k)$

k and R are as in the previous section, and let $A = M_2(k)$. Let V be a two-dimensional vector space over k, so that $A = M_2(k) \cong \mathrm{End}(V)$.

Our first task is to describe the complete R-lattices in V. It is a result in commutative algebra (Steinitz) that if M is a finitely generated, torsion-free R-module (where R is a Dedekind domain), then $M \cong R^k \oplus I$ for some non-negative $k \in \mathbb{Z}$ and I an ideal of R. We will give a more concrete special case of this.

Proposition 269. *Any complete R-lattice of V is of the form*

$$L = Ix + Ry$$

where x, y form a basis of V, and I is a fractional ideal of R.

In particular, this tells us the orders in a quadratic number field are all of the form $\mathbb{Z} + \mathfrak{a}\mathbb{Z}$, where $\mathfrak{a}$ is some ideal. In a principal ideal domain such as the Gaussian integers, this gives the classification of orders: $\mathbb{Z} + fi\mathbb{Z}$ for positive $f \in \mathbb{Z}$.

Proof. (Theorem 2.2.9, Machlachlan and Reid, rephrased)

First, observe that we may assume without loss of generality that, with regards to any fixed basis e_1, e_2 of V, $L \subset Re_1 + Re_2$. For, there always exists some $d \in R$ such that $dL \subset Re_1 + Re_2$. Although we don't use this assumption in the proof, it may be helpful to the reader to keep in mind.

Let $y \in V$, $y \neq 0$. Write

$$I_y = \{\alpha \in k : \alpha y \in L\}.$$

In other words, $L \cap ky = I_y y$. Note that I_y is an R-module. Furthermore, it is a fractional ideal, since, as observed at the beginning of the proof, $dL \subset Re_1 + Re_2$ for some $d \in R$.

Let e_1, e_2 be an R-basis of V. We'll define I (this will be the I in the theorem statement) by:

$$I = \{\alpha \in k : \alpha e_1 \in L + ke_2\}.$$

Again, I is an R-module. Furthermore, it is a fractional ideal, since, as observed at the beginning of the proof, $dL \subset Re_1 + Re_2$ for some $d \in R$.

Our first task is to define x and y for which $L = Ix + I_y y$. Choosing $x = e_1$ and $y = e_2$ may not work simply because it may fail that $Ix \subset L$ (draw a picture of Ii in the case $R = \mathbb{Z}$ to understand the issue). Therefore, we will have to adjust slightly, to $x = e_1 - \gamma e_2$; what follows is the precise way to do that.

Write

$$1 = \sum \alpha_i \beta_i$$

for some $\alpha_i \in I$ and $\beta_i \in I^{-1}$ (possible since $II^{-1} = R$). Write

$$\alpha_i e_1 = \ell_i + \gamma_i e_2 \quad \ell_i \in L, \gamma_i \in k,$$

for each i. We obtain

$$e_1 = \ell + \gamma e_2, \quad \ell = \sum \beta_i \ell_i, \;\; \gamma = \sum \beta_i \gamma_i \in k.$$

We now choose $x = e_1 - \gamma e_2$ and $y = e_2$.

The next goal is to show that $L = Ix + I_y y$. First, note that

$$Ix = I(e_1 - \gamma e_2) = I\left(\sum b_i \ell_i\right) \subset L.$$

For the first inclusion, we have

$$Ix + I_y y = I(e_1 - \gamma e_2) + L \cap ky \subset L$$

For the second inclusion, write $z \in L$ as $z = \alpha(e_1 + \beta e_2) \in L$. Then $\alpha e_1 \in L + ke_2$, so $\alpha \in I$. Then

$$z - \alpha x = \alpha(\beta + \gamma)e_2 \in L \cap ky = I_y y.$$

Finally, we we need only find some non-zero $y' \in V$ such that $I_{y'} = R$. Suppose we already have $L = Ix + I_y y$. Choose δ_1 and $\delta_2 \in k$ with

$$\delta_1 I^{-1} + \delta_2 I_y^{-1} = R.$$

(Any two ideals have some k-linear combination that is R.) Now I claim the appropriate choice of y' is $y' = \delta_1 x + \delta_2 y$. For,

$$I_{y'} = (I\delta_1^{-1}) \cap (I_y \delta_2^{-1}) = (\delta_1 I^{-1} + \delta_2 I_y^{-1})^{-1} = R.$$

(For the first equality, write out the meanings of the sets carefully; for the second, note that $(A+B)^{-1} = A^{-1} \cap B^{-1}$ for fractional ideals in general, since $A^{-1} = \{\alpha \in k : \alpha A \subset R\}$.) $\square$

Proposition 270. *$M_2(R)$ is a maximal order in $M_2(k)$.*

Proof. Suppose not. Then there is some order $\mathcal{O}$ which is strictly larger, hence containing some matrix with entries not in R. Performing row and column operations with matrices from $M_2(R)$, we obtain an element of the form

$$\alpha = \begin{pmatrix} a & 0 \\ 0 & 1 \end{pmatrix}, \alpha^n = \begin{pmatrix} a^n & 0 \\ 0 & 1 \end{pmatrix}, a \in k \setminus R.$$

Therefore $R[a]$ is not an R-lattice. But it is isomorphic to $R[\alpha]$, which is a submodule of the R-lattice $\mathcal{O}$, a contradiction. $\square$

For any complete R-lattice L, define

$$\mathrm{End}(L) = \{\sigma \in \mathrm{End}(V) : \sigma L \subset L\}.$$

If e_1, e_2 is a basis allowing us to identify $\mathrm{End}(V)$ with $M_2(k)$, then the maximal order $M_2(R) = \mathrm{End}(L)$ for $L = Re_1 + Re_2$. In particular, any conjugate of $M_2(R)$ is also a maximal order.

Proposition 271. *Let L be a complete R-lattice. Then* $\mathrm{End}(L)$ *is an order of $M_2(k)$. Further, every order is contained in some order of this form.*

Proof. Exercise. $\square$

Using the expression $L = Ix + Ry$, we obtain a nice form for these orders. They are the conjugates of sets of this form:

$$M_2(R; I) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} : a, d \in R, b \in I, c \in I^{-1} \right\}.$$

In the case that R is a principal ideal domain, then $M_2(R; I)$ is conjugate to $M_2(R)$.

9.10 Quaternion Algebras and Quadratic Forms

In this section, let A be a quaternion algebra over a field F. The norm map on A is a quadratic form on this space, so A is a quadratic space, i.e. a vector space with quadratic form. This also gives a symmetric bilinear form. The pure quaternions A_0 form a three-dimensional quadratic subspace.

This *norm form* is:

$$x_1^2 - ax_2^2 - bx_3^2 + abx_4^2$$

which is non-degenerate.

The standard basis $1, i, j, ij$ forms an orthogonal basis.

A quadratic space with quadratic form q is *isotropic* if there are non-zero vectors v with $q(v) = 0$.

Recall that a quaternion algebra may either be a division algebra or $M_2(F)$; the latter case is called *splitting* (A splits over F).

Theorem 272. *Suppose $A = \left(\frac{a,b}{F}\right)$. The following are equivalent.*

1. *A splits*
2. *A is isotropic*
3. *A_0 is isotropic*
4. *$ax^2 + by^2 = 1$ has a solution x, y over F.*
5. *a is in the image of the norm map from $F(\sqrt{b})$ down to F*

Proof. (Slight reshuffle of Theorem 2.3.1 in Maclachlan and Reid) **(1)** $\iff$ **(2)** Failing to be a division algebra is equivalent to having a non-zero non-invertible element, which must have norm 0.

(2) $\implies$ (3) Computation in Maclachlan and Reid. Exercise: find something more compelling.

(3) $\implies$ (4) The fact that $n(x) = 0$ for some $x = xi + yj + zij$ is the statement that

$$-ax^2 - by^2 + abz^2 = 0.$$

This is only possible if at most one of x, y, z is zero. If z is non-zero, this becomes

$$a(y/az)^2 + b(x/bz)^2 = 1.$$

If $z = 0$, we can write

$$a((1+a)/2a)^2 + b(x(1-a)/2ay)^2 = 1.$$

(4) $\implies$ (5) Rearrange the equation $ax^2 + by^2 = 1$ to be

$$N_{F(\sqrt{b})/F}(1/x + \sqrt{b}y/x) = a.$$

If $x = 0$, then $F(\sqrt{b}) = F$ and the norm map is the identity.

(5) $\implies$ (1) We exhibit a zero divisor. If $F(\sqrt{b}) = F$, then $\sqrt{b} + j$ is a zero divisor since $(\sqrt{b} + j)(\sqrt{b} - j) = b^2 - b^2 = 0$. Otherwise, by (5), we have

$$a = x^2 - by^2$$

In particular, $n(x + i + yj) = x^2 - a - by^2 = 0.$ ☐

Theorem 273. *Let $A = \left(\frac{a,b}{F}\right)$ and $C = \left(\frac{c,d}{F}\right)$ be quaternion algebras. The following are equivalent:*

1. *A and C are isomorphic*
2. *A_0 and C_0 are isometric quadratic spaces*
3. *$ax^2 + by^2 - abz^2$ and $cx^2 + dy^2 - cdz^2$ are equivalent quadratic forms over F*

Proof. (Remix of Maclachlan and Reid, Theorems 2.3.4 and 2.3.5) **(1) $\implies$ (2)** The pure quaternions are characterised as those x not in the centre whose squares x^2 are in the centre. Therefore an algebra isomorphism takes pure quaternions to pure quaternions. Furthermore, for pure quaternions, $n(x) = -x^2$. Therefore norm is preserved by the isomorphism also.

(2) $\implies$ (1) Suppose A_0 and C_0 are isometric by $\phi : A_0 \to C_0$. We wish to show that $1, \phi(i), \phi(j), \phi(i)\phi(j)$ forms a standard basis for C. That is, we show their squares are a, b and $-ab$, that all three are pure quaternions, and that the form an orthogonal basis.

Since ϕ preserves norms, it preserves squares, so

$$\phi(i)^2 = a, \quad \phi(j)^2 = b.$$

In A_0, i and j are orthogonal: $ij + ji = 0$. Therefore $\phi(i)$ and $\phi(j)$ are orthogonal:

$$\phi(i)\phi(j) + \phi(j)\phi(i) = 0$$

Orthogonality of the other pairs can be checked from this. Therefore,

$$(\phi(i)\phi(j))^2 = -ab.$$

One also concludes that $\phi(i)\phi(j)$ does not commute with $\phi(i)$ hence is not in the centre, while its square is. Therefore $\phi(i)\phi(j) \in C_0$. Finally, we must show $\phi(i)$, $\phi(j)$ and $\phi(i)\phi(j)$ are independent. Suppose

$$\alpha\phi(i) + \beta\phi(j) + \gamma\phi(i)\phi(j) = 0$$

Multiply by $\phi(i)$ on the left:

$$a\alpha + \beta\phi(i)\phi(j) - a\gamma\phi(j) = 0$$

This implies $\alpha = 0$. Similarly $\beta = \gamma = 0$.

(2) $\iff$ **(3)** Via the norm form, these are two ways of saying the same thing. □

thm:ASO **Theorem 274.** *There is an exact sequence*

$$1 \longrightarrow Z(A^*) \longrightarrow A^* \longrightarrow SO(A_0, n; F) \longrightarrow 1$$

For the proof we will need a little linear algebra. Let V, q be a quadratic space (with bilinear form $\langle \cdot, \cdot \rangle$). If $\mathbf{v}$ is an anisotropic vector, then we can define a reflection

$$\tau_{\mathbf{v}}(\mathbf{x}) = \mathbf{x} - \frac{2\langle \mathbf{x}, \mathbf{v} \rangle}{q(\mathbf{v})}\mathbf{v}.$$

This reflects in a hyperplane orthogonal to $\mathbf{v}$. In other words, it fixes vectors orthogonal to $\mathbf{v}$ and satisfies $\tau_{\mathbf{v}}(\mathbf{v}) = -\mathbf{v}$. It is a general fact that these reflections generate the orthogonal group of V.

Let us consider this situation for the pure quaternions in a quaternion algebra, with the norm form. Here,

$$n(x) = -x^2, \quad \langle x, y \rangle = \frac{1}{2}(xy + yx).$$

Therefore,

$$\tau_y(x) = x - \frac{xy + yx}{y^2}y = -yxy^{-1}.$$

Proof. The map
$$A^* \to O(A_0, n), \quad a \mapsto (x \mapsto axa^{-1})$$
is clearly well-defined. Its kernel is clearly $Z(A^*)$. It remains to show that the image is $SO(A_0, n)$. Note that on any anisotropic $a \in A_0^*$, the map has the form

$$a \mapsto (x \mapsto -\tau_a(x)).$$

In particular, the images of A_0^* have determinant 1, as negatives of reflections. Now, all products of these generate $SO_2(A_0, n; F)$, hence the image contains $SO_2(A_0, n; F)$.

Suppose the image were larger: then it would be all of $O(A_0, n; F)$. Then $-Id$ would be the image of some a, i.e.
$$-x = axa^{-1}$$
As $(-Id)^2 = Id$, we must have $a^2 \in Z(A^*)$, hence a is a pure quaternion. Taking $x = a$, we obtain $-a = a$, a contradiction. □

As a corollary, consider $A = M_2(\mathbb{R})$. Then

$$M_2(\mathbb{R})^* = \mathrm{GL}_2(\mathbb{R})$$

with center $\{kI : k \in \mathbb{R}^*\}$. Furthermore, we need to check the signature of the norm form on pure elements. The pure elements are those of trace zero; the determinant form on these looks like $-a^2 - bc$, which is of signature $(1, 2)$. Therefore,

$$\mathrm{PGL}_2(\mathbb{R}) \cong SO(2, 1; \mathbb{R}).$$

9.11 Quaternion Algebras over local fields

By a local field, I mean a p-adic/$\mathfrak{p}$-adic field or $\mathbb{R}$.

The symbol $\left(\frac{a,b}{F}\right)$ is called a Hilbert symbol and often defined, when F is a local field (p-adic or $\mathbb{R}$) in this way:

$$(a,b)_F := \left(\frac{a,b}{F}\right) = \begin{cases} 1 & ax^2 + by^2 = 1 \text{ has a solution in } F \\ -1 & ax^2 + by^2 = 1 \text{ has no solution in } F \end{cases}$$

The first is the split case, i.e. $(a,b)_F = 1$ if and only the quaternion algebra $\left(\frac{a,b}{F}\right)$ is isomorphic to $M_2(F)$.

Proposition 275. *The following hold:*

1. $(a, b^2)_F = 1$
2. $(a, b)_F = (b, a)_F$

3. $(a, 1-a)_F = 1$ *(provided both are in F^*)*

Proof. These are all pretty immediate from the definition above. $\square$

It turns out that using the symbols 1 and -1 is meaningful in the sense that $(a, b_1 b_2)_F = (a, b_1)_F (a, b_2)_F$. We don't prove that here. There is also a famous product formula:

$$\prod_v (a, b)_v = 1$$

where the product is taken over all places of $\mathbb{Q}$, i.e. $\mathbb{R}$ and all p-adic fields. This is equivalent to quadratic reciprocity! Exciting things lay down that path, but we aren't headed there now. See Borevich and Shafarevich.

Theorem 276. *The quaternion algebras over $\mathbb{R}$ are, up to isomorphism, only two: Hamilton's quaternions $\left(\frac{-1,-1}{\mathbb{R}}\right)$ and $M_2(\mathbb{R})$.*

Proof. We have already seen that the quaternion algebras over a field depend on a and b only up to squares; in fact, the possibilities are only

$$(a, b) = (-1, -1), (1, -1), (-1, 1), (1, 1).$$

The first is Hamilton's quaternions and the others are $M_2(\mathbb{R})$. $\square$

Now let us consider $\mathfrak{p}$-adic fields, i.e. our set-up is this (aligning with Maclachlan and Reid except I use $\mathfrak{p}$ for $\mathcal{P}$):

- K a $\mathfrak{p}$-adic field
- R its ring of integers
- $\mathfrak{p}$ its unique maximal ideal
- π a uniformizer, i.e. $\mathfrak{p} = (\pi)$
- $\overline{K}$ the residue field $R/\mathfrak{p}$, which is finite
- ν the discrete valuation associated to the uniformizer

Suppose A is a division algebra.

Proposition 277. *The map*

$$w : A \to \mathbb{Z}, \quad w(x) = \begin{cases} \nu(n(x)) & x \neq 0 \\ \infty & x = 0 \end{cases}$$

is a valuation on A.

Proof. See Maclachlan and Reid. ☐

Therefore, it has a valuation ring

$$\mathcal{O} = \{x \in A : w(x) \geq 0\}$$

and ideal

$$\mathcal{Q} = \{x \in A : w(x) > 0\}.$$

Theorem 278. *Let $F = K(\sqrt{u})$ be the unique unramified quadrati extension of K. The quaternion algebras over K are, up to isomorphism, only two: $\left(\frac{u,\pi}{K}\right)$ and $M_2(K)$.*

Proof. See Maclachlan and Reid. ☐

Don't forget we also have:

Theorem 279. *The quaternion algebras over $\mathbb{C}$ are all isomorphic to $M_2(\mathbb{C})$.*

9.12 Quaternion algebras over number fields

Let k be a number field. Let v be a valuation, either archimedean or non-archimedean. Let k_v be the completion. The various valuations of a number field are called *places* and they are associated to:

1. The real embeddings of k
2. The conjugate pairs of complex embeddings of k
3. The prime ideals of the ring of integers of k.

Then a quaternion algebra A over k gives rise to a quaternion algebra

$$A_v := A \otimes_k k_v$$

over k_v. If $A_v \cong M_2(k_v)$, then we say A *splits*. Otherwise it is *ramified*.

The analogy here is that a degree n extension of number fields L/K is an algebra over K; if a prime $\mathfrak{p}$ (with associated valuation v) of K splits in L, then $L \otimes_K K_v \cong K_v^n$ as a K_v-algebra. Otherwise it is some other product of finite extensions of K_v (with, in some sense, fewer zero divisors).

The principal local-to-global result is the following:

Theorem 280. *Let A be a quaternion algebra over a number field k. Then A splits over k if and only if $A \otimes_k k_v$ splits for all places v.*

See Maclachlan and Reid for more here.

10 Trace Fields and Quaternion Algebras for Kleinian Groups

We will be interested in the case of Kleinian Γ which lift to $\widehat{\Gamma} < \mathrm{SL}_2(\mathbb{C})$. When is this lifting possible? The sequence

$$1 \to \{\pm I\} \to \mathrm{SL}_2(\mathbb{C}) \to \mathrm{PSL}_2(\mathbb{C}) \to 1$$

is not split. For example (due to Kra, *On Lifting Kleinian Groups to* $\mathrm{SL}(2, \mathbb{C})$), an elliptic element $g : z \mapsto e^{2\pi ik/n}z$ of order n has two potential lifts:

$$\pm \begin{pmatrix} e^{\pi ik/n} & 0 \\ 0 & e^{-\pi ik/n} \end{pmatrix}$$

but when n is even, both of these have order $2n$, hence we cannot lift the group generated by g. Culler (*Lifting Representations to Covering Groups*, Advances in Mathematics) proves the following:

Theorem 281. *Let Γ be a finitely generated Kleinian group with no two-torsion. Then Γ lifts.*

Definition 282. *Let Γ be a non-elementary subgroup of* $\mathrm{PSL}_2(\mathbb{C})$. *Let $\widehat{\Gamma}$ be its preimage in* $\mathrm{SL}_2(\mathbb{C})$. *Then the* trace field *of Γ is*

$$\mathbb{Q}(tr\ \Gamma) = \mathbb{Q}(tr\ \gamma : \gamma \in \widehat{\Gamma}),$$

and, if Γ lifts to $\mathrm{SL}_2(\mathbb{C})$, *then, considering* $\Gamma < \mathrm{SL}_2(\mathbb{C})$, *the* quaternion algebra *of Γ is*

$$A_0\Gamma = \left\{ \sum a_i\gamma_i : a_i \in \mathbb{Q}(tr\ \Gamma), \gamma_i \in \Gamma,\ \textit{sum is finite} \right\}.$$

Note that the traces of the two primages of a matrix of $\mathrm{PSL}_2(\mathbb{C})$ differ only in sign.

Note that we need to lift to $\mathrm{SL}_2(\mathbb{C})$ for $A_0\Gamma$ because we expect this algebra to tensor up to $M_2(\mathbb{C})$, preserving the matrix structure (more on this later).

Let us consider the case of the Bianchi group $\mathrm{PSL}_2(\mathcal{O}_K)$. Here the trace field is clearly K. The quaternion algebra is not defined, at least if the group has two-torsion. But we will see later that we can move to a finite index subgroup to resolve this.

Our goal is to show that in general the trace field is a number field and the quaternion algebra is a quaternion algebra over that number field.

10.1 Trace Fields

thm:tracenumber

Theorem 283. *If Γ is Kleinian of finite covolume, with no two-torsion, then the trace field is a number field (i.e. finite extension of $\mathbb{Q}$).*

We will need this result:

Theorem 284. *Any finitely generated Kleinian group is finitely presented.*

Finitely presented means it can be given by finitely many generators with finitely many relations. We will not prove this; it follows from the *Scott core theorem.* We will also need the following.

Theorem 285 (Mostow-Prasad global rigidity)**.** *Let Γ_1 and Γ_2 be isomorphic finite covolume Kleinian groups. Let $\phi : \Gamma_1 \to \Gamma_2$ be the isomorphism. Then there exists some hyperbolic isometry g so that*

$$\phi(x) = gxg^{-1}.$$

A consequence of this is that a compact orientable 3-manifold has at most one unique hyperbolic structure of finite volume.

Proof. (Remix of Maclachlan and Reid and also Macbeath (*Commensurability of Co-Compact Three-Dimensional Hyperbolic Groups*)) Let Γ be as in the statement. Then it is geometrically finite and so finitely generated and finitely presented. Choose a finite presentation with generators $g_1, \ldots, g_n$. Also, it is isomorphic to a subgroup of $\mathrm{SL}_2(\mathbb{C})$.

Let

$$\mathrm{Hom}(\Gamma, \mathrm{SL}_2(\mathbb{C})) = \{\text{homomorphisms } \Gamma \to \mathrm{SL}_2(\mathbb{C})\}.$$

Let $\rho \in \mathrm{Hom}(\Gamma, \mathrm{SL}_2(\mathbb{C}))$ be injective.

Since $\rho(\Gamma)$ is non-elementary (a consequence of finite covolume), we may assume that it contains loxodromic elements g and h such that $\langle g, h\rangle$ is irreducible. Since conjugation does not change the trace, we may assume that g fixes 0 and ∞ and h fixes 1. Note that Γ has a finite presentation with generators g, h, g_1, …, g_n.

Now, let V be the subvariety of $\mathrm{Hom}(\Gamma, \mathrm{SL}_2(\mathbb{C}))$ which is defined by mapping the generators $g, h, g_1, \ldots, g_n$ of Γ into $\mathrm{SL}_2(\mathbb{C})$ with this fixed point normalization.

I claim V is an algebraic set over $\mathbb{Q}$ (i.e. it is cut out by finitely many equations over $\mathbb{Q}$). To see this, consider the entries of the images of the finitely many generators as the variables. Then the equation are given by the determinant 1 condition, the relations on Γ, and the fixed point equations above.

Our goal will be to prove that the dimension of this algebraic set V is zero. For, if so, then it must consist of finitely many points, with algebraic coordinates. Therefore the entries of the elements of Γ in any $\mathrm{SL}_2(\mathbb{C})$ representation are algebraic, up to conjugation. Then there is a finite extension K of $\mathbb{Q}$ containing them all. Since the trace field is a subfield of K, it is a number field. Finally, the trace field is invariant under conjugation.

If the dimension of V is non-zero, then there would be elements of V arbitrarily close to ρ. Local rigidity[1] says that for $\rho' \in V$ sufficiently close to ρ, $\rho'(\Gamma)$ is also of finite covolume and isomorphic to $\rho(\Gamma)$. By Mostow-Prasad rigidity, $\rho(\Gamma)$ and $\rho'(\Gamma)$ are conjugate by a

[1] Suppose Γ is Kleinian of finite covolume. Suppose $\rho : \Gamma \to \mathrm{SL}_2(\mathbb{C})$ is an injection. Then any $\rho' : \Gamma \to \mathrm{SL}_2(\mathbb{C})$ sufficiently close to ρ is an isomorphism to its image and $\rho(\Gamma)$ has finite covolume. See Maclachlan and Reid, Theorem 1.6.2.

hyperbolic isometry. There are only finitely many such conjugations that preserve the condition that g fixes 0 and ∞ and h fixes 1 (since it must permute these fixed points). Therefore any ρ' sufficiently close to ρ satisfies $\rho' = \rho$. Therefore the dimension must be zero.

□

10.2 Quaternion Algebras

Theorem 286. *Let $\Gamma < \mathrm{SL}_2(\mathbb{C})$ be non-elementary. Then $A_0\Gamma$ is a quaternion algebra over the trace field. More precisely, if g and h are loxodromic elements such that $\langle g, h\rangle$ is irreducible, then 1, g, h, gh form a basis.*

Proof. It is a $\mathbb{Q}(tr\ \Gamma)$-algebra by definition. Let g and h be loxodromic with $\langle g, h\rangle$ irreducible. Recall that $\langle g, h\rangle$ is irreducible if and only if $tr\ [g, h] \neq 2$. Now, if we place the matrices I, g, h, gh as columns in a matrix $M(g, h)$, then

$$\det M(g, h) = 2 - tr\ [g, h].$$

Therefore, I, g, h, gh are linearly independent over $\mathbb{C}$. In particular, then, $A_0\Gamma \otimes_{\mathbb{Q}(tr\ \Gamma)} \mathbb{C} \cong M_2(\mathbb{C})$. Since $M_2(\mathbb{C})$ is a quaternion algebra over $\mathbb{C}$, $A_0\Gamma$ is a quaternion algebra over $\mathbb{Q}(tr\ \Gamma)$. □

As a consequence of the proof, the quaternion trace and norm are the usual matrix trace and determinant.

Theorem 287. *If Γ is a non-elementary subgroup of $\mathrm{SL}_2(\mathbb{C})$, with loxodromic g and h such that $\langle g, h\rangle$ is irreducible, then Γ is conjugate to a subgroup of $\mathrm{SL}_2(K)$ where K is a quadratic extension of $\mathbb{Q}(tr\ \Gamma)$.*

Proof. We can conjugate so that

$$g = \begin{pmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{pmatrix}, \quad h = \begin{pmatrix} a & 1 \\ c & d \end{pmatrix}$$

where $c \neq 0$, $\lambda \neq 1$. Here, λ is quadratic over $\mathbb{Q}(tr\ \Gamma)$, generating some field k over $\mathbb{Q}(tr\ \Gamma)$. Since $a + d$ and $a\lambda + d\lambda^{-1}$ are in $\mathbb{Q}(tr\ \Gamma)$, it must be that a and d are in k. Therefore $c = ad - 1 \in k$. □

In particular, if the trace field has real embeddings, then Γ can be conjugated into $\mathrm{SL}_2(\mathbb{R})$. This is because a loxodromic g with real trace is hyperbolic, hence has real eigenvalues.

10.3 Invariant trace field and invariant quaternion algebra

Maclachlan and Reid's motivational example is the figure-eight knot complement group

$$\Gamma = \left\langle \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ -w & 1 \end{pmatrix} \right\rangle$$

clearly has trace field $\mathbb{Q}(\sqrt{-3})$. However, by adding a new generator

$$\begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}$$

we obtain a group Δ such that $[\Delta : \Gamma] = 2$ (to see this, check that the new matrix normalizes Γ and is of order two). But now the trace field is strictly larger, as it contains i (it is not hard to find an example, e.g. take the product of the matrices above in the order opposite that which they appear above).

The point here being that the trace field is not an invariant of the commensurability class of Γ. The idea is to define a sort of 'minimal' trace field for a commensurability class.

Definition 288. *Suppose* $\Gamma < \mathrm{SL}_2(\mathbb{C})$ *is finitely generated.*
Define $\Gamma^{(2)} := \langle \gamma^2 : \gamma \in \Gamma \rangle$.
The invariant trace field *of* Γ *is* $k\Gamma := \mathbb{Q}(tr\ \Gamma^{(2)})$.
The invariant quaternion algebra *of* Γ *is* $A\Gamma := A_0\Gamma^{(2)}$.

Note that $\Gamma^{(2)}$ is normal in Γ, and, since Γ is finitely generated, its quotient is an elementary abelian 2-group.

Theorem 289. *The invariant trace field and invariant quaternion algebra are invariants of the commensurability class of* Γ.

Proof. (From Maclachlan and Reid)

Claim: $\mathbb{Q}(tr\ \Gamma^{(2)})$ **is contained in the trace field of any finite index subgroup of** Γ**.**

This claim will suffice to show that $k\Gamma$ is a commensurability invariant. For, suppose Δ and Γ are commensurable. Then $\Gamma^{(2)}$ and $\Delta^{(2)}$ are commensurable and their intersection Λ is finite index in both. By definition, $k\Gamma$ and $k\Delta$ contain $\mathbb{Q}(tr\ \Lambda)$. But if the claim is proven, then $k\Gamma$ and $k\Delta$ are both contained in $\mathbb{Q}(tr\ \Lambda)$ also. Therefore $k\Gamma = \mathbb{Q}(tr\ \Lambda) = k\Delta$.

Proof of Claim: Suppose $\Delta < \Gamma$ is of finite index. Then

$$C = \bigcap_{g \in \Gamma} g \Delta g^{-1}$$

is of finite index (since Δ is of finite index, we need only take the intersection over finitely many coset representatives) and normal. If we can prove the claim for C, then it is proven for Δ. Therefore we may assume without loss of generality that Δ is normal.

We will show that $\Gamma^{(2)} \subset A_0\Delta$, from which the result will follow. Let $g \in \Gamma$. Since Δ is normal, g induces an automorphism by conjugation. Therefore it induces an automorphism of $A_0\Delta$, extending linearly. But, as $A_0\Delta$ is a quaternion algebra, all automorphisms are inner (Skolem Noether Theorem). Therefore there exists some invertible $a \in A_0\Delta$ so that

$$gxg^{-1} = axa^{-1}, \ \ \forall x \in A_0\Delta.$$

In other words, $g^{-1}a$ is in the centre of $A_0\Delta \otimes_{\mathbb{Q}(tr\ \Delta)} \mathbb{C}$, i.e. $g^{-1}a = yI$ for some $y \in \mathbb{C}$. Therefore $g^2 = y^{-2}a^2$. It therefore suffices to show that $y^2 \in \mathbb{Q}(tr\ \Delta)$:

$$y^2 I = \det(g^{-1}a)I = \det(a)I = a^2 - tr\ (a)a \in A_0\Delta.$$

This concludes the proof of the claim.

Now we must show that $A\Gamma$ is a commensurability invariant. The invariant quaternion algebra is generated over $k\Gamma$ by a pair of loxodromic elements g and h such that $\langle g, h\rangle$ is irreducible. But if Γ and Δ are commensurable, then $\Gamma^{(2)} \cap \Delta^{(2)}$ is of finite index in both. If we may choose g and h from this intersection, which shows $A\Gamma = A\Delta$. But loxodromic elements are non-torsion, so we may take suitable powers g^n, h^n of g and h which lie in $\Gamma^{(2)} \cap \Delta^{(2)}$. And if $\langle g, h\rangle$ is irreducible, then $\langle g^n, h^n\rangle$ is irreducible (use the characterisation of irreducibility in terms of sharing a fixed point). □

Some observations:

1. If Γ is of finite covolume, then its invariant trace field is a number field. If it were real, then $\Gamma^{(2)}$ would be conjugate to a subgroup of $\mathrm{SL}_2(\mathbb{R})$, contradicting finite covolume. So **the invariant trace field of a finite covolume Kleinian group cannot have a real embedding.**

2. If Γ is non-elementary with a parabolic g (e.g. finite covolume but not co-compact), then $g - I$ is non-invertible in the quaternion algebra, so $A_0\Gamma$ is split. So **the quaternion algebra of a cofinite non-cocompact group is split**.

3. These invariants of the commensurability class are *not* complete invariants.

Finally, we have a map

$$\Gamma \to SO((A\Gamma)_0, n)$$

given by

$$(g \in \Gamma) \mapsto (x \mapsto gxg^{-1}).$$

That is, conjugation by g induces an automorphism of $\Gamma^{(2)}$, hence $A\Gamma$. It preserves the norm on $(A\Gamma)_0$. We use Theorem 274.

10.4 The Trace

Some basic facts about traces.

prop:tracefacts **Proposition 290.** *Let $X, Y \in \mathrm{SL}_2(\mathbb{C})$.*

1. *There are monic polynomials $p_n(x), q_n(x) \in \mathbb{Z}[x]$, of degree $n-1$ and $n-2$ respectively, such that*
$$X^n = p_n(tr\ X)X - q_n(tr\ X)I.$$
In particular, $X^2 = (tr\ X)X - I$.

item:monictrace

2. *$tr\ X^n$ is a monic integral polynomial in $tr\ X$, of degree n. For example, $tr\ X^2 = (tr\ X)^2 - 2I$.*

3. *Trace is invariant on conjugacy classes.*

4. *Trace of a product is invariant under cyclic permutation of the factors. For example, $tr\ XY = tr\ YX$.*

5. *$tr\ X = tr\ X^{-1}$*

6. *$tr\ XY + tr\ XY^{-1} = (tr\ X)(tr\ Y)$*

Exercise 291. *Verify all this.*

Here's some more, just to give an idea:

Proposition 292. 1. *$tr\ [X, Y] = (tr\ X)^2 + (tr\ Y)^2 + (tr\ XY)^2 - tr\ Xtr\ Ytr\ XY - 2I$*

2. *$tr\ XYXZ = tr\ XYtr\ XZ - tr\ YZ^{-1}$*

3. *$tr\ XYX^{-1}Z = tr\ XYtr\ X^{-1}Z - tr\ X^2YZ^{-1}$*

4. *$tr\ X^2YZ = tr\ Xtr\ XYZ - tr\ YZ$*

5. *$tr\ XYZ + tr\ YXZ + tr\ Xtr\ Ytr\ Z = tr\ Xtr\ YZ + tr\ Ytr\ XZ + tr\ Ztr\ XY$*

6. *$2tr\ XYZW = tr\ Xtr\ YZW + tr\ Ytr\ ZWX + tr\ Ztr\ WXY + tr\ Wtr\ XYZ + tr\ XYtr\ ZW - tr\ XZtr\ YW + tr\ XWtr\ YZ - tr\ Xtr\ Ytr\ ZW - tr\ Ytr\ Ztr\ XW - tr\ Xtr\ Wtr\ YZ - tr\ Ztr\ Wtr\ XY + tr\ Xtr\ Ytr\ Ztr\ W$*

7. *The reader is referred to Maclachlan and Reid for more, e.g. for $tr\ XYZtr\ X'Y'Z'$ etc.*

Proof. None of this is interesting in and of itself. See Maclachlan and Reid, Section 3.4. □

Machlachlan and Reid use the above, and lots of induction, to give the following generators for trace fields.

tracegenerators **Theorem 293.** *Let $\Gamma < \mathrm{SL}_2(\mathbb{C})$ be generated by $g_1, \ldots, g_n$. Then*

1. *$\mathbb{Q}(tr\ \Gamma)$ is generated by*
$$\{tr\ g_i\} \cup \{tr\ g_i g_j\} \cup \{tr\ g_i g_j g_k\}$$

2. *$k\Gamma$ is genreated by*
$$\{tr\ g_i^2\} \cup \{tr\ g_i^2 g_j^2\} \cup \{tr\ g_i^2 g_j^2 g_k^2\}$$

Some particular consequences of this are:

1. If $\Gamma = \langle g, h \rangle$ is non-elementary, g and h not of order two, then
$$\begin{aligned} \mathbb{Q}(tr\ \Gamma) &= \mathbb{Q}(tr\ g, tr\ h, tr\ gh) \\ k\Gamma &= \mathbb{Q}(tr\ ^2g, tr\ ^2h, tr\ g\ tr\ h\ tr\ gh) \end{aligned}$$

2. If $\Gamma = \langle f, g, h \rangle$, then
$$\mathbb{Q}(tr\ \Gamma) = \mathbb{Q}(tr\ f, tr\ g, tr\ h, tr\ fg, tr\ fh, tr\ gh, tr\ fgh)$$

There are many more in Machlachlan and Reid, Section 3.5.

We also establish a generalization of Theorem 283, wherein we can eliminate the need for no two-torsion.

hm:tracenumber2 **Theorem 294.** *If Γ is Kleinian of finite covolume, then the trace field is a number field.*

Proof. Since Γ is finitely generated, Selberg's Lemma applies, and we may let Γ' be a torsion free, finite index, subgroup. By Theorem 283, Γ' has finite degree trace field. But there is an integer n such that for any $g \in \Gamma$, we have $g^n \in \Gamma'$. Therefore, by Proposition 290 part (2), every trace in Γ is of degree at most n over $\mathbb{Q}(tr\ \Gamma')$. Therefore every trace is algebraic. By Theorem 293, we are done. $\square$

The next goal is to give a standard basis for the quaternion algebra. We have shown that if $\langle g, h \rangle$ is irreducible, then $1, g, h, gh$ is a basis, but we haven't checked it is standard; it may not be. Standard means that i^2 and j^2 are in $k\Gamma$, and $ij = -ji$. This is the same as saying i and j are orthogonal under the norm form; we can rephrase what we need as saying i, j, ij form an orthogonal basis of pure quaternions. An appropriate basis is found by:

1. translating g and h so they are pure quaternions, i.e.
$$g' = g - \frac{tr\ g}{2} I, \quad h' = h - \frac{tr\ h}{2} I$$

2. applying Gram Schmidt to obtain

$$h'' = h' - \frac{B(g', h')}{B(g', g')} g'$$

Then, finding the norms of g' and h'', we prove the first part of the following.

Theorem 295. *Let Γ be Kleinian, with elements g, h not of order two, with g not parabolic, and such that $\langle g, h \rangle$ is irreducible. Then*

$$A\Gamma = \left(\frac{tr\ ^2 g\ (tr\ ^2 g - 4), tr\ ^2 g\ tr\ ^2 h\ (tr\ [g,h] - 2)}{k\Gamma} \right).$$

Furthermore, if in addition $g, h \in \Gamma^{(2)}$, then we can simplify to

$$A\Gamma = \left(\frac{tr\ ^2 g - 4, tr\ [g,h] - 2}{k\Gamma} \right).$$

10.5 Examples

First, consider the Bianchi groups $\Gamma = \mathrm{PSL}_2(\mathcal{O}_K)$. These are of finite covolume with the number of cusps equal to the class number, i.e. they have parabolics. Therefore, $A\Gamma = M_2(k\Gamma)$.

Suppose $\alpha \in \mathcal{O}_K$. Then

$$\begin{aligned}
&\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ \alpha & 1 \end{pmatrix} \in \Gamma \\
\implies &\begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 2\alpha & 1 \end{pmatrix} \in \Gamma^{(2)} \\
\implies &\begin{pmatrix} 1+4\alpha & 2 \\ 2\alpha & 1 \end{pmatrix} \in \Gamma^{(2)} \\
\implies &2 + 4\alpha \in k\Gamma \\
\implies &\alpha \in k\Gamma
\end{aligned}$$

Therefore, $k\Gamma = K$.

See also the Figure 8 knot complement and the compact tetrahedron in Machlachlan and Reid.

11 Applications and places to go

11.1 Discreteness criterion

Here's a first example of the relevance of the invariant trace field. Taken from Maclachlan and Reid, Section 5.1.

Theorem 296. *Let* Γ *be a finitely generated subgroup of* $\mathrm{PSL}_2(\mathbb{C})$ *which satisfies:*

1. *tr* (Γ) *consists of algebraic integers.*
2. $\Gamma^{(2)}$ *is irreducible.*
3. *For each embedding* $\sigma : k\Gamma \to \mathbb{C}$*, the image* $\sigma(tr\ (\Gamma^{(2)})$ *is bounded.*

Then Γ *is discrete.*

Proof. Since Γ is finitely generated, $k\Gamma$ is a number field.

12 Errata

12.1 Maclachlan and Reid

1. Theorem 1.6.2 should say 'is an isomorphism to its image' instead of 'is an isomorphism'.
2. Lemma 2.8.2 and its preamble ought to say that $\mathrm{End}_A(M)$ is to be considered a *right* A-module.
3. Proof of Theorem 2.1.8. This calls on 2.1.7. But it actually calls on the statement that a 4-dimensional simple algebra is either a division algebra or $M_2(F)$; as stated it is assuming what is to be proved.
4. Proof of Theorem 3.3.4, last line: probably it was meant that $\Gamma^{(2)} \subset A_0\Gamma_1$ instead of $\Gamma^{(2)} \subset A_0(\Gamma)$.