NOTES FOR PATTERSON–SULLIVAN MEASURES ON CIRCLES

MXM FALL 2024

Abstract. We introduce the background and motivation for our project. This includes, in order, group actions on spaces, the hyperbolic plane and its boundary, convex projective domains in the projective plane, and Patterson–Sullivan measures.

Things will be presented at a level of rigor and precision somewhere between what one would see in a math textbook or paper and what one might expect in a popular science book. The aim is to get the reader up to speed somewhat quickly and hopefully without too much pain.

0. A note on topology

We won't do any topology, but there will be some topology lurking in the background, some we make some comments about it here.

A topology on a set X is, roughly speaking, a structure that gives us a way of describing which points are close together which is less quantitative / "squishier" than a metric.

This structure works by distinguishing certain subsets of X, conventionally called the open sets of X, which can be thought of as generalizing open balls in $\mathbb{R}$. (There are axioms on the set of open sets which generalize things that happen with open balls in $\mathbb{R}$.) Two points of X can be distinguished by the topology if there is an open set of X containing one but not the other.

For example, $\mathbb{R}$ has a topology where the open sets are generated by the open balls (under unions and finite intersections). In this topology, every pair of points is topologically distinguishable. If we zoom in far enough, there is no difference, topologically speaking, between the interval $[0, 1]$ and the interval $[0, 0.00001]$ (one difference between the structure of the topology and the structure of the metric space ...)

Another example: $\mathbb{Z}$ has a topology where every subset is an open set.

1. Group actions

1.1. Groups.

A *group* is an algebraic structure that was designed to capture the structure of sets of symmetries.

For instance: think of a square. There are eight symmetries preserving the square, which we list here.

- Do nothing.
- Clockwise rotation by $\frac{\pi}{2}$, π, or $\frac{3\pi}{2}$.
- Reflection across the vertical or horizontal midlines.
- Reflection across one of the two diagonals.

(If we're being particular: here we imagine putting the square into the Euclidean plane—the 2-dimensional geometry we're familiar with—and by "symmetries" we mean maps from the square to itself which preserve Euclidean distances and angles.)

We can compose these symmetries to get other symmetries, there is a "do nothing" symmetry, and given any one of these symmetries there is some other symmetry we can compose it with to end up with the "do nothing" symmetry. The definition of a group generalizes this set-up. Below, the operation $\cdot$ should be thought of as composition of symmetries.

Definition 1.1. A *group* is a set G with an operation $\cdot\colon G \times G \to G$ such that

- $g_1 \cdot (g_2 \cdot g_3) = (g_1 \cdot g_2) \cdot g_3$,
- there exists $e \in G$ such that $g \cdot e = e \cdot g = g$ for all $g \in G$, and
- for each $g \in G$, there exists $g^{-1} \in G$ such that $g \cdot g^{-1} = g^{-1} \cdot g = e$.

Example 1.2. The **dihedral group of order 8** is the set of symmetries preserving the square described above, where $\cdot$ can be defined by an explicit 8-by-8 "multiplication table".

Exercise 1.3. Find this multiplication table.

Example 1.4. The set of integers $\mathbb{Z}$ is a group, where the group operation is given by addition.

Example 1.5. The set of integer n-tuples $\mathbb{Z}^n = \{(z_1, \ldots, z_n) : z_1, \ldots, z_n \in \mathbb{Z}\}$ is a group, where the group operation is given by coordinate-wise addition.

We can describe the group using the *group presentation*

$$\langle z_1, \ldots, z_n \mid z_i \cdot z_j = z_j \cdot z_i \text{ for } 1 \leq i, j \leq n \rangle.$$

This means there are group elements $z_1, \ldots, z_n$ (we can think of z_i as the element which is 1 in the i^{th} coordinate and 0 elsewhere), called the *generators*, such that every group element can be represented as a combination of the generators in some order, and all of the redundancies are represented by the *relations* (which in this case, tell us that the group operation on a pair of generators in either order produces the same output).

Exercise 1.6. The *free group on n generators*, denoted F_n, is the group given by the group presentation $\langle g_1, \ldots, g_n \rangle$ (with no relations).

The exercise is to parse what this means.

Example 1.7. Let $\mathrm{SL}(2, \mathbb{R})$ be the set of 2-by-2 real matrices with determinant 1. This is a group where the operation $\cdot$ is given by matrix multiplication.

Example 1.8. Let $\mathrm{SL}(2, \mathbb{Z})$ be the set of 2-by-2 integer matrices with determinant 1. This is also a group where the operation $\cdot$ is given by matrix multiplication. It is a *subgroup* of $\mathrm{SL}(2, \mathbb{R})$.

1.2. **Group homomorphisms.** A *homomorphism* between two groups is a map (i.e. a function) between them that respects the group structures. Concretely, this means the following.

Definition 1.9. Given two groups G and H, a map $\phi\colon G \to H$ is a *homomorphism* if

$$\phi(g_1 \cdot g_2) = \phi(g_1) \cdot \phi(g_2)$$

for all $g_1, g_2 \in G$. Note that we are doing group operations in G on the left-hand side, and group operations in H on the right-hand side.

A homomorphism between G and itself will also be called a homomorphism on G.

Example 1.10. $z \mapsto -z$ is a homomorphism on $\mathbb{Z}$.

So is $z \mapsto 0$.

Example 1.11. Given any group G and any $g \in G$, the map $c_g \colon G \to G$ defined by $c_g(h) = ghg^{-1}$ is a homomorphism on G.

Exercise 1.12. Find a homomorphism from the dihedral group of order 8 to $\mathrm{SL}(2, \mathbb{R})$.

Homomorphisms allow us to define a precise notion of when two groups "have the same structure". Such groups are really the same structure, just perhaps with different names.

Definition 1.13. Given two groups G and H, the homomorphism $\phi \colon G \to H$ is an *isomorphism* if it is one-to-one and onto.

G and H are *isomorphic* if there exists an isomorphism between them.

Example 1.14. c_g is an isomorphism from G to itself.

Example 1.15. Consider the groups

$$S^1 := \{z \in \mathbb{C} : |z| = 1\} = \{e^{i\theta} : \theta \in \mathbb{R}\}$$

where the group operation is multiplication of complex numbers, and

$$\mathrm{SO}(2) := \left\{ R_\theta := \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} : \theta \in \mathbb{R} \right\}$$

where the group operation is matrix multiplication.

There is an isomorphism from S^1 and $\mathrm{SO}(2)$ given by $e^{i\theta} \mapsto R_\theta$.

1.3. **Group actions.** Given a group G and a set S, a *group action* of G on S should be thought of as "how G shows up as a set of symmetries of S."

Definition 1.16. Given a group G and a set S, an *action* (more precisely, a left action) of G on S is a map $\varphi \colon G \times S \to S$ such that

- $\varphi(e, s) = s$ for all $s \in S$, and
- $\varphi(h, \varphi(g, s)) = \varphi(h \cdot g, s)$ (where $\cdot$ denotes the group operation in G) for all $s \in S$ and all $g, h \in G$.

Here, you should think of $\varphi(g, s)$ as "where g takes s". Note the conditions above imply that for each fixed g, the map from S to itself given by $\cdot \mapsto \varphi(g, \cdot)$ has $\cdot \mapsto \varphi(g^{-1}, \cdot)$ as an inverse, and in particular must be bijective. This leads us to the following equivalent characterization.

Definition 1.17. We write $\mathrm{Bij}(S)$ to denote the set of bijections from S to itself.

Given a group G and a set S, an *action* of G on S is a group homomorphism $\phi \colon G \to \mathrm{Bij}(S)$.

To restate the equivalence between the two definitions:

- given an action map $\varphi \colon G \times S \to S$ (per Definition 1.16), $g \mapsto (\cdot \mapsto \varphi(g, \cdot))$ is a group homomorphism from G to $\mathrm{Bij}(S)$;

- given a group homomorphism $\phi\colon G \to \mathrm{Bij}(S)$, $\varphi(g,s) = \phi(g)(s)$ is an action map.

Example 1.18. $\mathrm{SL}(2,\mathbb{R}) \subset \mathrm{Bij}(\mathbb{R}^2)$, so a homomorphism from the dihedral group of order 8 to $\mathrm{SL}(2,\mathbb{R})$ gives us a group action of this dihedral group on $\mathbb{R}^2$.

Example 1.19. There is an action of $\mathbb{Z}^2 = \langle a, b : ab = ba \rangle$ on $\mathbb{R}^2$ where a acts as $(x,y) \mapsto (x+1,y)$ and b acts as $(x,y+1)$.

More generally, given a basis $(\vec{v_1}, \ldots, \vec{v_n})$ of $\mathbb{R}^n$, there is an action of

$$\mathbb{Z}^n = \langle z_1, \ldots, z_n \mid z_i z_j = z_j z_i \text{ for } 1 \le i,j \le n \rangle$$

on $\mathbb{R}^n$ where z_i acts as $\vec{v} \mapsto \vec{v} + \vec{v_i}$.

Example 1.20. Consider the subgroup $\mathrm{Isom}^+(\mathbb{E}^2)$ of $\mathrm{Bij}(\mathbb{R}^2)$ which preserves the usual Euclidean metric [and orientation] in the plane. The inclusion of this subgroup into $\mathrm{Bij}(\mathbb{R}^2)$ gives us a group action of this group on the plane.

(We can also more concretely identify $\mathrm{Isom}^+(\mathbb{E}^2)$ as a semidirect product of two 2-by-2 matrix groups ... but we will not worry more about this.)

This group action is *transitive*: given any two points $p, q \in \mathbb{R}^2$, there exists $g \in \mathrm{Isom}^+(\mathbb{E}^2)$ such that $g(p) = q$.

One working definition of a "geometry" (usually attributed to Felix Klein and described as part of his Erlangen program) is a pair (X, G) where X is a set and G acts transitively on X. From this point of view, a geometry is characterized by its symmetries.

1.4. Discrete group actions.

The group actions in Example 1.20 and Example 1.19 are both actions on $\mathbb{R}^2$, but there is one qualitative difference between the group action in Example 1.20 and the one in Example 1.19.

Definition 1.21. Suppose we have a group Γ acting on a set X. Given a point $x \in X$, the set of points

$$\Gamma \cdot x := \{\gamma \cdot x : \gamma \in \Gamma\}$$

is called the Γ-*orbit of* x.

If we consider a(ny) point $o \in \mathbb{R}^2$, the orbit $\mathbb{Z}^2 \cdot o$ is a discrete set: there exist a set of pairwise disjoint open balls around each of these points. The same is not true for the $\mathrm{Isom}^+(\mathbb{E}^2)$-orbit (which by transitivity is all of $\mathbb{R}^2$).

Indeed, $\mathbb{Z}^n$ is an example of a discrete group. A group is discrete if it does not contain any sequence of elements which accumulate to a group element.[1] Examples of non-discrete groups are given by $\mathbb{R}^n$ (with addition defined coordinate-wise), where e.g. $\epsilon\vec{e_1} \to \vec{0}$ as $\epsilon \to 0$, or $\mathrm{SL}(2,\mathbb{R})$, where e.g. $\begin{bmatrix} e^{\epsilon} & \\ & e^{-\epsilon} \end{bmatrix} \to I$ as $\epsilon \to 0$. We will not discuss enough topology to define this notion of accumulation more precise, and make just two further remarks on this for now.

Remark 1.22. Any finitely-generated group (i.e. any group with a group presentation with finitely many generators) is discrete.

Remark 1.23. Any discrete group acting on a metric space by isometries must have discrete orbits.

[1] This assumes we have a topology on the group. We will sweep this under the carpet: in all the examples we encounter, there will be a natural (or conventional) topology.

Further references.

- Lyons, Introduction to Groups and Geometries, Chapter 2: Groups
- Lyons, Introduction to Groups and Geometries, Chapter 3.1: Geometries and models

2. The hyperbolic plane

Hyperbolic geometry is a non-Euclidean geometry. It was first discovered by mathematicians who were trying to prove that Euclid's fifth postulate[2] follows from the first four postulates for Euclidean geometry. In fact, it does not, and there are geometries where the first four postulates hold but the fifth does not.

Spherical geometry is one of them: on (the surface of) the sphere, great circles are analogue of straight lines, and given any great circle C and any point p not on it, there is no great circle through p that does not intersect C.

Hyperbolic geometry is another: in hyperbolic space, given a geodesic (the analogue of a straight line) γ and a point p not on it, there are infinitely many geodesics through p that do not intersect γ.

There are many (equivalent) ways of thinking about / working with hyperbolic geometry. We will introduce four. The first two will be useful for computation and visualization. We will use the third mostly as a way to get to the fourth, which in turn will be motivation for the next section.

We will also stick to dimension 2 here, although all of the ideas can be generalized to all dimensions.

2.1. The upper half-plane model.

Consider the upper half-plane

$$\mathbb{H} := \{x + iy \in \mathbb{C} : y > 0\}$$

(where $i^2 = -1$) with the metric arclength element $ds = \frac{\sqrt{dx^2+dy^2}}{y}$.

Whoa wait what? The metric is a device that allows us to measure distances and angles. We will not attempt to explain it carefully, or work with it directly. We will just remark that $dx^2 + dy^2$ is the usual Euclidean arclength element / gives us the usual Euclidean metric, and what we have here is the Euclidean metric scaled by something inverse to the square of the height.

This arclength element is like a ruler. If we imagine an object of constant size moving around $\mathbb{H}$, the closer the object gets to the real axis, the smaller it will look to us with our Euclidean eyes. What looks like one unit according to the Euclidean metric is actually $\frac{1}{y^2}$ units according to the hyperbolic metric, and this blows up as $y \to 0$; in other words, 1 unit in the hyperbolic metric will look like y^2 units according to the Euclidean metric, which shrinks towards 0 as $y \to 0$. Conversely, the further the object gets from the real axis, the bigger it will look to us with our Euclidean eyes.

Here is an illustration of this courtesy of Escher + a conformal mapping[3]: all of the white tiles in that image are the same size according to the hyperbolic metric, as are all of the black tiles!

---

[2] This is about the uniqueness of parallel lines: given any [bi-infinite] straight line ℓ in Euclidean space and a point p not on it, there is a unique line through p that does not intersect ℓ.

[3] Escher created a version of this inside a circle; some complex analysis + code was used to map it into the upper half-plane. See the next subsection!

Okay so how do we work with the metric? Instead of working directly with the metric, we will interact with the space by thinking about geodesics and isometries.

Geodesics are distance-minimizing paths; they are the analogues of straight lines in the Euclidean plane.

In the upper half-plane model $\mathbb{H}$, the geodesics are vertical (straight) lines and semicircles which are perpendicular to the real axis $\{y = 0\}$. This can be taken as a black box; if you wish to think about why it is true, one possible approach (which can be made a proof with more elbow grease) is presented in the series of exercises at the end of this subsection.

$\mathrm{SL}(2, \mathbb{R})$ acts on $\mathbb{H}$ as fractional linear transformations, meaning

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

Exercise 2.1. Prove that this is, in fact, a group action.

In fact, $\mathrm{SL}(2, \mathbb{R})$ preserves the metric; it is basically the group of isometries of $\mathbb{H}$. To remove the word "basically", we make two observations.

(1) Every element in $\mathrm{SL}(2, \mathbb{R})$ preserves orientation.[4]
(2) Notice that flipping the signs of all the entries produces another matrix which acts on $\mathbb{H}$ as the same fractional linear transformation. In other words, the matrices A and $-A$ in $\mathrm{SL}(2, \mathbb{R})$ act the same way, so we may as well look at the action of the quotient group $\mathrm{PSL}(2, \mathbb{R}) := \mathrm{SL}(2, \mathbb{R})/\pm I$, i.e. the group obtained from $\mathrm{SL}(2, \mathbb{R})$ by identifying every pair of matrices $(A, -A)$.

Thus: the group of *orientation-preserving* isometries of $\mathbb{H}$ is given by $\mathrm{PSL}(2, \mathbb{R})$.

The boundary. We define the boundary $\partial\mathbb{H}$ to be the real axis $\mathbb{R} = \{y = 0\} \subset \mathbb{C}$, together with a point at infinity ∞. Sequences in $\mathbb{H}$ converge towards points in the boundary in a "what you see is what you get" way:

- $z_n = x_n + iy_n \to \infty$ if and only if $y_n \to \infty$
- $z_n = x_n + iy_n \to x \in \mathbb{R}$ if and only if $y_n \to 0$ and $x_n \to x$.

With this notion of convergence, there is a natural way to identify a geodesic ℓ with the pair of points in $\partial\mathbb{H}$ given by the endpoints of ℓ.

Proof sketch for identification of geodesics.

Exercise 2.2. Convince yourself that the imaginary half-axis $\{x = 0, y > 0\}$ is a geodesic (i.e. a distance-minimizing path between any two points on it.)

Exercise 2.3. Show that the action of $\mathrm{PSL}(2, \mathbb{R})$ on $\mathbb{H}$ by fractional linear transformations extends naturally to $\partial\mathbb{H}$. Moreover, this action is triply transitive, meaning for any two triples $(x, y, z), (x', y', z') \in (\partial\mathbb{H})^3$, there exists $g \in \mathrm{SL}(2, \mathbb{R})$ such that $(gx, gy, gz) = (gx', gy', gz')$.

Exercise 2.4. Let L be the set of vertical (straight) lines in $\mathbb{H}$ and semicircles in $\mathbb{H}$ which are perpendicular to the real axis $\{y = 0\}$.

Show that the action of $\mathrm{PSL}(2, \mathbb{R})$ on $\mathbb{H}$ by fractional linear transformations takes elements of L to elements of L.

[4] Informally: if we draw a tiny cat in $\mathbb{H}$, the action of $\mathrm{SL}(2, \mathbb{R})$ will send it to another tiny cat that is *not mirrored*.

2.2. **The Poincaré disk model.** Consider the unit disk

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

with the metric $ds = \frac{2\sqrt{dx^2+dy^2}}{1-(x^2+y^2)}$. We call this the Poincaré disk model.

As with the upper half-plane model, we will not work directly with this metric. We will note that this has the same qualitative behavior as the upper half-space model: as we approach the unit circle $\{|z| = 1\}$, the metric element ("ruler") ds blows up relative to the Euclidean metric element $\sqrt{dx^2 + dy^2}$ (since $x^2 + y^2$ approaches 1). Hence if there are two objects of the same (hyperbolic) size, one at the origin $0 \in \mathbb{C}$ and one closer to the unit circle, the former will look larger than the latter.

Here is an illustration of that due to Escher.

We can map conformally (= in an angle-preserving way) between the upper half-plane and the unit disk, thanks to the magic of complex analysis: concretely, the map

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

maps $\mathbb{H}$ to $\mathbb{D}$, taking i to 0 and the positive imaginary axis to the interval $(-1, 1)$. This allows us to use properties of the upper half-plane to obtain properties of the Poincaré disk (and vice versa).

Geodesics and isometries. Geodesics in $\mathbb{D}$ are given by semicircular arcs perpendicular to the unit circle $\{z \in \mathbb{C} : |z| = 1\}$.

Isometries of $\mathbb{D}$ can be obtained (for instance) by translating the action of $\mathrm{SL}(2, \mathbb{R})$ on $\mathbb{H}$ by fractional linear transformations through the conformal mapping $\mathbb{H} \to \mathbb{D}$. This identifies the isometries of $\mathbb{D}$ as a different matrix group which is isomorphic to $\mathrm{PSL}(2, \mathbb{R})$.

The boundary. We can define $\partial\mathbb{D}$ to be the unit circle $\{z \in \mathbb{C} : |z| = 1\}$. As in the case of $\mathbb{H}$ and $\partial\mathbb{H}$, sequences and geodesics in $\mathbb{D}$ converge to points in $\partial\mathbb{D}$ in a "what you see is what you get" way.

2.3. **The hyperboloid model.** Consider the subset $\{x^2 + y^2 - z^2 = -1\}$ of $\{(x, y, z) : x, y, z \in \mathbb{R}\} = \mathbb{R}^3$. This is a 2-sheeted hyperboloid; take the sheet S^+ with $z > 0$. We can define a metric on this using the bilinear form associated to the quadratic form $(x, y, z) \mapsto x^2 + y^2 - z^2$, but we will not work with this.

The Poincaré disk can be obtained from the hyperboloid model by projecting S^+ through a point on the other sheet of the hyperboloid onto a perpendicular plane.

In the hyperboloid model, geodesics are given by intersections of S^+ with planes through the origin.

The orientation-preserving isometries can be shown to be given by $\mathrm{SO}^+(1, 2)$, the group of 3-by-3 matrices of determinant one which preserve the bilinear form associated to quadratic form $(x, y, z) \mapsto x^2 + y^2 - z^2$ and preserve the sign of the first coordinate.

(If you are wondering why this is not the same group: there is a group isomorphism identifying $\mathrm{PSL}(2, \mathbb{R})$ with $\mathrm{SO}^+(1, 2)$ [possibly modulo a small lie.])

2.4. **The projective disk model.** Projecting S^+ through the origin $(0,0,0) \in \mathbb{R}^3$ onto a perpendicular plane gives us a different disk model of the hyperbolic plane (sometimes called the Beltrami–Klein model.)

In this model, the geodesics are straight lines (in the usual Euclidean sense) and the orientation-preserving isometry group is $\mathrm{SO}^+(1,2)$.

2.5. **Discrete group actions on the hyperbolic plane.** Earlier, we saw an action of $\mathbb{Z}^2$ on the plane with discrete orbits (see Section 1.4.) Here we consider some discrete subgroups of $\mathrm{PSL}(2,\mathbb{R})$: these will produce group actions on $\mathbb{H}$ with discrete orbits.

We start with subgroups generated by single elements. These are examples of *elementary* subgroups: each of their orbits in $\mathbb{H}$ has a small number of accumulation points (at most 2) on the boundary $\partial\mathbb{H}$.

Proposition 2.5. *Any $g \in \mathrm{PSL}(2,\mathbb{R})$ is one of the following:*

- elliptic*, meaning g fixes a unique point in $\mathbb{H}$;*
- parabolic*, meaning g fixes a unique point in $\partial\mathbb{H}$; or*
- hyperbolic*, meaning g fixes a geodesic ℓ in $\mathbb{H}$ (as a set, not pointwise) and translates along ℓ. The fixed geodesic in this case is called the* translation axis *of g.*

Moreover, g is elliptic if and only if $|\operatorname{tr} g| < 2$, parabolic if and only if $|\operatorname{tr} g| = 2$, and hyperbolic otherwise.

Exercise 2.6. Prove the proposition! (Hint: given $g \in \mathrm{PSL}(2,\mathbb{R})$, look at the points of $\mathbb{H} \cup \partial\mathbb{H}$ fixed by the action of g.)

Remark 2.7. We can also distinguish between the three classes using the *translation length* of g, which is defined as

$$\tau(g) := \inf_{z \in \mathbb{H}} d_{\mathbb{H}}(z, g \cdot g).$$

If this infimum is realized, $\tau(g)$, the smallest distance g moves any point in $\mathbb{H}$.

g is elliptic if and only if $\tau(g) = 0$, and there exists $z \in \mathbb{H}$ for which $d_{\mathbb{H}}(z, g \cdot z) = 0$.

g is parabolic if and only if $\tau(g) = 0$, but the infimum is not realized.

g is hyperbolic otherwise, i.e. if and only if $\tau(g) > 0$; in this case the infimum is always realized, along the translation axis of g

Example 2.8. A cyclic group generated by a single parabolic or hyperbolic element is a discrete subgroup of $\mathrm{PSL}(2,\mathbb{R})$ isomorphic to $\mathbb{Z}$.

It acts on $\mathbb{H}$ with discrete orbits.

More complicated examples are given by free groups and surface groups.

Example 2.9. Pick two hyperbolic elements g, h whose translation axes have distinct endpoints $g^\pm, h^\pm$ in $\partial\mathbb{H}$, and such that there exist disjoint neighborhoods $A^\pm$ of $g^\pm$ and $B^\pm$ of $h^\pm$ such that

- $\mathbb{H} - (A^+ \cup A^- \cup B^+ \cup B^-) \neq \varnothing$,
- $g^{\pm 1} \cdot (\mathbb{H} - A^\mp) \subset A^\pm$ and
- $h^{\pm 1} \cdot (\mathbb{H} - B^\mp) \subset B^\pm$.

Then $\langle g, h \rangle$ is a subgroup of $\mathrm{PSL}_2(\mathbb{R})$ isomorphic to F_2, the free group on two generators, by the following ("ping-pong") argument: if there is any relation between g and h, then some word in g and h gives the identity element, and so must fix every point in $\mathbb{H}$. But any word in g and h takes any point $z \in \mathbb{H} - (A^+ \cup A^- \cup B^+ - \cup B^-)$ into one of A^+, A^-, B^+, B^-. In particular, no word in g and h can fix such a z. Hence are no relations, and g and h generate a free group.

Examining the ping-pong argument more closely also shows: such a free subgroup of $\mathrm{PSL}_2(\mathbb{R})$ acts on $\mathbb{H}$ with discrete orbits.

This last example will be described less precisely, at least for now: computing with examples like it is (possible, but takes some complex analysis + serious hyperbolic trigonometry, and so) a bit of a pain, so we won't start there ...

Example 2.10. In Example 1.19, if we take $\mathbb{R}^2$ and identify each $\mathbb{Z}^2$-orbit to a single point, we obtain a quotient space $\mathbb{R}^2/\mathbb{Z}^2$ isomorphic to a torus (the surface of a donut). We can reverse-engineer this example by starting with a square, whose opposite sides can be identified to obtained a torus, giving it a Euclidean metric / identifying it as a patch of the Euclidean plane, and finding group elements acting on the Euclidean plane as our two translations.

The torus is a genus-1 surface (the sphere is a genus-0 surface). We can create surfaces of genus $g \geq 2$ by taking a $4g$-gon (e.g. for $g = 2$, an octagon) and identifying pairs of sides appropriately. When we do this, all of the vertices of the $4g$-gon are identified to a single point in the quotient. In $\mathbb{H}$, we can find a regular $4g$-gon whose internal angles add up to 2π, and elements of $\mathrm{PSL}_2(\mathbb{R})$ which identify pairs of sides in the way we want This allows us to identify the $4g$-gon with that patch of $\mathbb{H}$ and the quotient of $\mathbb{H}$ by the group Γ generated by these elements of $\mathrm{PSL}_2(\mathbb{R})$ with the surface of genus g.

This group action Γ also has discrete orbits: informally, the $4g$-gon (in $\mathbb{H}$) and its translates by elements of Γ do not overlap (except along edges), and there is one orbit point inside each translate of the $4g$-gon.

Further references.

- Lyons, Introduction to Groups and Geometries, Chapter 3.2: Möbius geometry
- Lyons, Introduction to Groups and Geometries, Chapter 3.3: Hyperbolic geometry
- Mangahas, The Ping-Pong Lemma, from *Office Hours with a Geometric Group Theorist*

3. Convex projective domains

The projective disk model can be generalized in a way that we will now describe. A very rough description of this might go: the projective disk model was an especially symmetric blob in the plane with straight line geodesics. We can now draw slightly less symmetric blobs in the plane and declare the geodesics to be straight lines in this blob.

3.1. **The real projective plane.** A more precise description, which will actually allow us to give some structure and get some grip on these "blobs", starts with the real projective plane $\mathbb{R}\mathbf{P}^2$. As a space: $\mathbb{R}\mathbf{P}^2$ is the set of all lines through the origin in $\mathbb{R}^3$. Equivalently, $\mathbb{R}\mathbf{P}^2$ can be thought of (and defined!) as the unit sphere (around the origin in $\mathbb{R}^3$), except with opposite points identified: every line through the origin in $\mathbb{R}^3$ passes through exactly one pair of opposite points on the unit sphere, and a pair of opposite points on the unit sphere exactly identifies one line through the origin.

This last description gives us a topology on $\mathbb{R}\mathbf{P}^2$ in terms of the topology on the unit sphere: a set of lines is close if they pass through nearby points on the unit sphere.

There are *homogeneous coordinates* on $\mathbb{R}\mathbf{P}^2$ given by $[x : y : z]$ where $(x, y, z) \neq (0, 0, 0)$, and $[rx : ry : rz]$ and $[x, y, z]$ are taken to represent the same point[5] for any $r \in \mathbb{R} \setminus \{0\}$.

The set of all points with homogeneous coordinate $x = 0$ is parametrized by the homogeneous coordinates $[y : z]$, which also parametrize $\mathbb{R}\mathbf{P}^1$, the set of lines through the origin in $\mathbb{R}^2$. Indeed, this is the set of lines which pass though the origin in $\mathbb{R}^3$ and lie in the yz-plane. Such a set of points can be described as a projective line.

The set of all points in $\mathbb{R}\mathbf{P}^2$ with homogeneous coordinate $x \neq 0$ is equivalently parametrized by $\{(y, z) : y, z \in \mathbb{R}\} = \mathbb{R}^2$.

What happened in the last two paragraphs can be generalized: a projective line in $\mathbb{R}\mathbf{P}^2$ can be described as the set of all lines through the origin in $\mathbb{R}^3$ which lie in some fixed plane π through the origin in $\mathbb{R}^3$.

The complement of any projective line is a subspace of $\mathbb{R}\mathbf{P}^2$ isomorphic to $\mathbb{R}^2$; we call such a subspace an *affine chart* of $\mathbb{R}\mathbf{P}^2$.

3.2. **Smooth convex domains and the Hilbert metric.** Consider an affine chart of $\mathbb{R}\mathbf{P}^2$. We have a notion of convexity in $\mathbb{R}^2$: an open subset $\Omega \subset \mathbb{R}^2$ is convex if for any two points $p, q \in \Omega$, the straight line segment from p to q lies entirely inside Ω. Since affine charts are isomorphic to $\mathbb{R}^2$, we have a notion of convexity in an affine chart.

Definition 3.1. A *properly convex domain* of $\mathbb{R}\mathbf{P}^2$ is an open subset $\Omega \subset \mathbb{R}\mathbf{P}^2$ such that $\overline{\Omega}$ is contained in an affine chart of $\mathbb{R}\mathbf{P}^2$, and Ω is convex in that affine chart.

Notice that a properly convex domain Ω naturally has a boundary $\partial\Omega$ (which can be defined more formally as $\overline{\Omega} \setminus \Omega$, but we will not worry too much about it here).

We say that a properly convex domain $\Omega \subset \mathbb{R}\mathbf{P}^2$ is *smooth* if its boundary contains no sharp points or proper line segments.

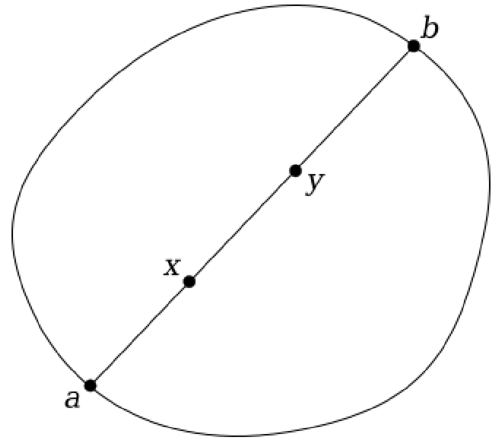
Definition 3.2. Given a properly convex domain Ω, we can define a metric d_Ω (called the Hilbert metric) on Ω as follows: to find $d_\Omega(x, y)$ draw the straight line segment between x and y: extend this to meet $\partial\Omega$ at a and b (so that a, x, y, b lie on the extended straight line segment in that order) and then define

$$d_\Omega(x, y) = \frac{1}{2} \ln \frac{|ay||xb|}{|ax||yb|}$$

where e.g. $|ay|$ denotes the Euclidean distance from a to y in the affine chart.

The quantity inside the logarithm is a quantity called the cross-ratio of the lines a, x, y, b (remember: points in $\mathbb{R}\mathbf{P}^2$ represent lines through the origin in $\mathbb{R}^3$). The point of this strange-looking definition is that the cross-ratio is the unique projective invariant of four lines: if we apply a symmetry g of $\mathbb{R}\mathbf{P}^2$ which takes our affine chart to another affine chart, and Ω to a different properly convex domain inside

[5]This does make it not exactly a coordinate system in the usual sense, but it is usual to preserve this quirk for reasons of symmetry ...



the new affine chart, the cross ratio of ga, gx, gy, gb is the same as the cross ratio of a, x, y, b, and so the Hilbert metric remains unchanged. (The logarithm ensures that distances satisfy the usual triangle inequality, and the $\frac{1}{2}$ is a normalizing factor chosen so that Example 3.4 works.)

The group $\mathrm{GL}(3, \mathbb{R})$ of all invertible 3-by-3 matrices (where the group operation is matrix multiplication) acts linearly on $\mathbb{R}^3$, and hence acts $\mathbb{R}\mathbf{P}^2$ since it takes lines through the origin in $\mathbb{R}^3$ to lines through the origin in $\mathbb{R}^3$.

The scalar diagonal matrices

$$\left\{ \begin{bmatrix} \lambda & & \\ & \lambda & \\ & & \lambda \end{bmatrix} : \lambda \in \mathbb{R} \smallsetminus \{0\} \right\} \subset \mathrm{GL}(3, \mathbb{R})$$

act trivially on $\mathbb{R}\mathbf{P}^2$ (i.e. every scalar diagonal matrix acts on $\mathbb{R}\mathbf{P}^2$ by fixing every point in $\mathbb{R}\mathbf{P}^2$, so if we consider the group $\mathrm{PGL}(3, \mathbb{R})$ obtained from $\mathrm{GL}(3, \mathbb{R})$ by identifying all scalar multiples, we have an action of $\mathrm{PGL}(3, \mathbb{R})$ on $\mathbb{R}\mathbf{P}^2$.

Definition 3.3. Define $\mathrm{Aut}(\Omega)$ to be the subgroup of $\mathrm{PGL}(3, \mathbb{R})$ which fixes Ω.

Example 3.4. The projective disk model of the hyperbolic 2-space gives an example where Ω is a round disk in an affine chart and $\mathrm{Aut}(\Omega) \cong \mathrm{SO}^+(1, 2)$.

More generally, $\mathrm{Aut}(\Omega)$ acts by isometries with respect to the Hilbert metric d_Ω on Ω, and when Ω is smooth $\mathrm{Aut}(\Omega)$ is the isometry group.

3.3. **(Linear) Coxeter groups.** A large class of examples of convex domains different from the round disk is given by *Coxeter groups*. Informally, these can be thought of as groups generated by reflections.

Definition 3.5. A *Coxeter group* W_S is a group with a presentation

$$\langle s_1, \dots, s_N \mid (s_i s_j)^{m_{ij}} = e \text{ for } 1 \leq i, j \leq N \rangle,$$

where $m_{ii} = 1$ and $m_{ij} = m_{ji} \in \{2, 3, \dots, \infty\}$ for $i \neq j$.

In particular, $s_1^2 = \cdots = s_N^2 = e$, i.e. the generators behave as we expect reflections to behave: doing the same reflection twice in a row produces the same result as doing nothing.

Suppose W_S is a Coxeter group with $|S| = N$. We can realize W_S as a group of reflections in $V \cong \mathbb{R}^N$. We describe this construction here: a little bit of care is needed, to make sure, roughly speaking, the reflections don't crash into each other. (To visualize things, it may be easier to stick to $N \leq 3$ at first.)

Definition 3.6. A linear transformation $R\colon V \to V$ is a *reflection* if $R^2 = 1$ and R has eigenvalue -1 with multiplicity 1.

Suppose we have a reflection R. Let $\vec{v}$ be an eigenvector for R with eigenvalue -1, and $\alpha\colon V \to \mathbb{R}$ be a linear functional whose kernel is the 1-eigenspace of R and $\alpha(\vec{v}) = 2$. Then

$$R(\vec{x}) = \vec{x} - \alpha(\vec{v})\vec{x}$$

for all $\vec{x} \in V$.

Given reflections $R_1, \dots, R_n$, we associate to each R_i the data of a (-1)-eigenvector $\vec{v_i} \in V$ and linear functional α_i as in the previous paragraph. We consider the subset $\tilde{\Delta} \subset V$ given by the intersection of all the half-spaces $\{\vec{v} \in V : \alpha_i \geq 0\}$ for $1 \leq i \leq n$.

Definition 3.7. An N-by-N matrix $A = (A_{ij})_{1 \leq i,j \leq N}$ is *weakly compatible* with W_S if

- $A_{ii} = 2$ for all $1 \leq i \leq N$;
- $A_{ij} = 0$ for all $i \neq j$ with $m_{ij} = 2$;
- $A_{ij} < 0$ for any $i \neq j$ with $m_{ij} \neq 2$;
- $A_{ij}A_{ji} = 4\cos^2\left(\frac{\pi}{m_{ij}}\right)$ for all $i \neq j$ with $2 < m_{ij} < \infty$.

Definition 3.8. Suppose we have reflections $R_1, \dots, R_n$ with associated eigenvectors $(\vec{v_1}, \dots, \vec{v_n}) =: v$ and linear functionals $(\alpha_1, \dots, \alpha_n)$. Suppose the N-by-N matrix $A = (A_{ij})_{1 \leq i,j \leq N}$ given by $A_{ij} = \alpha_i(\vec{v_j})$ is weakly compatible with W_S.

Then we say A is the *Cartan matrix* of (α, v)

Definition 3.9. We say the Cartan matrix A of (α, v) is *compatible* with W_S if $A_{ij}A_{ji} \geq 4$ for all $i \neq j$ with $m_{ij} = \infty$.

By a theorem of Vinberg, when the Cartan matrix of (α, v) is compatible, $\tilde{\Delta}$ has non-empty interior, and the translates of $\tilde{\Delta}$ by the group generated by the reflections $R_1, \dots, R_n$ tile some convex subset of V. Thus: projecting $V \cong \mathbb{R}^N$ to $\mathbb{R}\mathbf{P}^{N-1}$ (in dimension $N = 3$: projecting from 3-dimensional space to the unit sphere, and then pairs of antipodal points on that sphere), whenever we have a compatible Cartan matrix for a Coxeter group W_S, we can realize W_S as a group of reflections of $\mathbb{R}^{|S|}$ which give a tiling of a convex domain in $\mathbb{R}\mathbf{P}^{N-1}$. Such a group of reflections is sometimes called a *linear Coxeter group.*

We say a Coxeter group W_S is *irreducible* if we cannot decompose S into disjoint subsets S_1, S_2 such that $W_S \cong W_{S_1} \times W_{S_2}$.

All of the Coxeter groups we deal with will be large, meaning it admits a surjection onto a non-abelian free group on two generators, and hyperbolic, meaning roughly that it acts nicely on a space that behaves like hyperbolic space (more on this in the last section below.)

By results of Vinberg and Benoist: if a Coxeter group is irreducible and large, it can be realized as group of reflections which tile a *proper* convex domain. Moreover, if W_S is hyperbolic, the convex domain can be taken to be smooth.

Example 3.10. A Coxeter group W_S with $|S| = 3$ generators is called a triangle group. These are defined by three (possibly infinite) integers $\ell, m, n \in \{2, \dots, \infty\}$ (corresponding to m_{12}, m_{13}, m_{23} in the more general definition of a Coxeter group.)

The tile Δ (the projection of $\tilde{\Delta}$ from $\mathbb{R}^3$ to $\mathbb{R}\mathbf{P}^2$) can be taken to a triangle with interior angles $\frac{\pi}{\ell}, \frac{\pi}{m}, \frac{\pi}{n}$, and depending on the sum of these angles, the W_S-translates of Δ do one of three things:

- (if the sum of angles is greater than π) they tile all of $\mathbb{R}\mathbf{P}^2$, or in other words we get a tiling of a sphere;
- (if the sum of angles is equal to π) they tile all of an affine chart, so we get a tiling of an Euclidean plane; or
- (if the sum of angles is less than π) they tile a disk contained inside an affine chart, i.e. [the projective model of] a hyperbolic plane.

Further references.

- Lyons, Introduction to Groups and Geometries, Chapter 3.5: Projective geometry
- For Coxeter groups, Section 3 of this paper (far too terse as an introductory text, but useful as reference after you have an idea of what you're looking for.)
- The Wikipedia article on triangle groups has some nice pictures of the tilings.

4. Patterson–Sullivan measures

Suppose we have a discrete group Γ acting by isometries on a metric space X, where X is either the hyperbolic plane $\mathbb{H}$ (equivalently, the Poincaré disk $\mathbb{D}$) or a round convex projective domain Ω. Equivalently, we realize Γ as a discrete subgroup of $\mathrm{Isom}^+(\mathbb{H})$ or $\mathrm{Aut}(\Omega)$.

We can pick a point $o \in X$ and ask: how does the orbit $\Gamma \cdot o$ behave? In particular, where in the boundary do the orbit points tend to go?

A Patterson–Sullivan measure μ_o is one answer to this last question. This is a measure on the boundary ∂X, a device which tells us how large subsets[6] of ∂X are. Roughly speaking, $\mu_o(A)$ is given by the proportion of orbit points $\Gamma \cdot o$ which accumulates to A.

4.1. **The Patterson construction.** To actually build such a measure for: first define the Poincaré series

$$Q_{\Gamma,o}(s) = \sum_{\gamma \in \Gamma} e^{-s \cdot d(o,\gamma o)}.$$

This converges when s is large, and diverges when s is small. We define the *critical exponent of* $Q_{\Gamma,o}$ to be

$$\delta_\Gamma := \inf\{s \geq 0 : Q_{\Gamma,o}(s) < \infty\}$$

(by an argument involving the triangle inequality, this number does not depend on o: for a given group $\Gamma \curvearrowright X$, $Q_{\Gamma,o}$ and $Q_{\Gamma,p}$ have the same critical exponent.)

A priori, $Q_{\Gamma,o}(\delta_\Gamma)$ may converge or diverge. We will assume (this will be true for all the group actions we consider in this project) that it diverges.

[6] Not all subsets, just the ones which are measurable. Patterson–Sullivan measures are Borel, meaning the measurable sets can be described in terms of the open sets of ∂X. When ∂X is the circle, for example, these are exactly the sets generated (under countable unions, coutable intersections, and complements) by open intervals of the circle.

Then, for any $s > \delta_\Gamma$, we define the following weighted sum of Dirac measures

$$\mu_{o,s} := \frac{1}{Q_{\Gamma,o}(s)} \sum_{\gamma \in \Gamma} e^{-s \cdot d(o,\gamma o)} \mathcal{D}_{\gamma o}.$$

Here a Dirac measure $\mathcal{D}_x$ is a measure than assigns 1 to any (measurable) subset of X containing the point x, and 0 to any other (measurable) subset of X.

Finally, we take a limit in the space of probability measures on $X \cup \partial X$:

$$\mu_o := \lim_{s \searrow \delta_\Gamma} \mu_{.}$$

It is a natural-enough construction once we make sense of the technicalities, although some work is needed to check that it does make sense and has good properties. That work was first done by Patterson and Sullivan.

4.2. **The shadow lemma.** We may or may not work directly with this definition. We can also attempt to use of a few very useful properties that Patterson–Sullivan measures have.

Proposition 4.1. *Patterson–Sullivan measures are conformal densities, meaning: if x and y are two points in X, then the measures μ_x and μ_y are absolutely continuous, and*

$$\frac{d\mu_x}{d\mu_y}(\xi) = e^{-\delta_\Gamma \beta_\xi(x,y)}$$

where δ_Γ is the critical exponent of $\Gamma \curvearrowright X$ and β_ξ is the Busemann function based at ξ.

We won't define Busemann functions or tried to make sense of absolutely continuity or derivatives of measures for now: we will just parse the last statement as "the two measures are not too different from each other, and we can describe exactly how they relate in terms of geometric quantities relating x and y."

This has the following consequence, which will be most directly useful for us:

Definition 4.2. Given $x, y \in X$ and $r > 0$, we define the shadow $\mathcal{O}_r(x, y)$ to be the set of all points in ∂X which are the endpoints of geodesic rays starting at x and intersecting an open ball of radius r centered at y

In other words: $\mathcal{O}_r(x, y)$ is the shadow cast on the boundary ∂X by the ball of radius r around y, if we imagine light rays emanating from a point source at x.

Theorem 4.3 (Shadow lemma)**.** *For any $x \in X$, there exists $r_0 > 0$ such that: for all $r \geq r_0$, there exists $C > 0$ so that*

$$\frac{1}{C} e^{-\delta_\Gamma d(o,\gamma o)} \leq \mu_o(\mathcal{O}_r(o, \gamma o)) \leq C e^{-\delta_\Gamma d(o,\gamma o)}$$

for all $\gamma \in \Gamma$.

4.3. **Limit sets and conical points.**

Definition 4.4. The *limit set* Λ_Γ of Γ (realized as a subgroup of $\mathrm{Isom}^+(\mathbb{H})$ or $\mathrm{Aut}(\Omega)$) is the set of accumulation points in ∂X of some (equivalently[7], any) orbit $\Gamma \cdot o$.

The group (action) Γ is *elementary* if $|\Lambda_\Gamma| \leq 2$, and *non-elementary* otherwise.

[7]for the particular X's that we have

One image in the upper half-space model: imagine the orbit points are raindrops falling through $\mathbb{H}$ towards the ground $\partial\mathbb{H}$; the limit set shows where the raindrops land. We will assume all of our group actions are non-elementary.

Proposition 4.5. *For any $o \in X$, the Patterson–Sullivan measure μ_o is supported on Λ_Γ, meaning $\mu_o(A) > 0$ if and only if $A \cap \Lambda_\Gamma \neq \varnothing$.*

The shadow lemma is especially useful when we are dealing with points in the limit set which are in infinite sequences of nested shadows. We give a name to such points.

Definition 4.6. Given $\epsilon > 0$ and $o \in X$, a point $x \in \Lambda_\Gamma$ is called (ϵ, o)-*conical* if there exists an infinite sequence (γ_n) of distinct group elements such that $x \in \bigcap_n \mathcal{O}_\epsilon(o, \gamma_n o)$.

The following result (which takes quite a lot of work to show) states that, from the point of view of the Patterson–Sullivan measures, we essentially only see uniformly conical limit points.

Theorem 4.7. *For any $o \in X$, there exists $\epsilon > 0$ such that the set of (ϵ, o)-conical limit points has full μ_o-measure.*

Further references.

- Sullivan's original paper — would *not* recommend reading in full: maybe just the first couple of paragraphs, and then skim for the pictures (Figure 2 is an illustration of a shadow, and Figure 3 is ... fun to gawk at.)

5. A little geometric group theory

This section should be fleshed out a little.

5.1. Hyperbolic groups.

One property of $\mathbb{H}$ is that geodesic triangles are thin. We can look at metric spaces which satisfy just this property; it turns out that this is sufficient to capture many of the large-scale features of $\mathbb{H}$.

Definition 5.1. A metric space X with metric d is *geodesic* if for every $x, y \in X$, there is a path in X whose length is $d(x, y)$.

A geodesic metric space X is *Gromov-hyperbolic* if there exists $\delta > 0$ such that: for any triangle $\Delta(x, y, z)$ with geodesic sides, the side $\overline{xz}$ is contained in the union of the δ-neighborhoods of the other two sides $\overline{xy}$ and $\overline{yz}$.

Definition 5.2. A group Γ is *word-hyperbolic* if it acts isometrically and discretely on a Gromov-hyperbolic space with compact quotient.

Equivalently, Γ is word-hyperbolic if and only if any Cayley graph of Γ equipped with the word metric is Gromov-hyperbolic as a metric space

All of the groups we consider will be word-hyperbolic.

5.2. Automata for hyperbolic groups.

Given a finitely-generated group Γ, it is in general difficult to systematically enumerate all of the group elements.

For word-hyperbolic groups, there are finite-state automata which will do this for us in a computationally efficient way[8].

We will likely make use of these.

[8] At least, as efficient as we could hope for, given that the number of elements grows exponentially in word-length.