

Group Actions on Hyperbolic Spaces

Lecture Notes

Christopher H. Cashen

Contents

Lecture 1. Metric spaces, Trees, and Free groups	1
1.1. Motivation and metric spaces	1
1.2. Paths, length spaces, and geodesics	1
1.3. Graphs and trees	2
1.4. The Gromov product and the boundary of a tree	3
1.5. Cantor boundaries and compactification	4
1.6. Isometries of regular trees	5
1.7. The free group and its Cayley tree	6
1.8. Centralizers in the free group	7
1.9. Ping-pong on the boundary	7
1.10. Exercises	8
Lecture 2. The hyperbolic plane and surface groups	10
2.1. The Poincaré disk and its metric	10
2.2. The isometry group	11
2.3. Geodesics and the circle at infinity	13
2.4. Classification of isometries	14
2.5. Hyperbolic surfaces and surface groups	15
2.6. Exercises	17

LECTURE 1

Metric spaces, Trees, and Free groups

1.1. Motivation and metric spaces

Goal. Study groups G (usually finitely presented) via the geometry of their actions. Finitely presented groups are countable and discrete, so tools from the theory of topological groups are not directly useful. Instead we focus on making the group act on a metric space.

- If a group has a sufficiently nice action on a metric space, the geometry of that space influences the group's properties.
- Negative curvature (hyperbolicity) is a group property.
- Hyperbolic groups have a well-defined boundary at infinity.

$$\begin{array}{ccc} \text{dynamics on the boundary} & \longleftrightarrow & \text{geometry of the space} \\ & & \longleftrightarrow \text{algebraic structure of the group} \end{array}$$

DEFINITION 1.1.1. A *metric space* (X, d) consists of a set X and a function $d: X \times X \rightarrow \mathbb{R}$ satisfying, for all $x, y, z \in X$:

Positive definiteness: $d(x, y) \geq 0$, with equality if and only if $x = y$.

Symmetry: $d(x, y) = d(y, x)$

Triangle inequality: $d(x, z) \leq d(x, y) + d(y, z)$

The metric topology has as a basis the open balls:

$$B_r(x) := \{y \in X \mid d(x, y) < r\}$$

Also define the closed balls and open and closed neighborhoods of subsets:

$$\bar{B}_r(x) := \{y \in X \mid d(x, y) \leq r\} \quad N_r(Y) := \{x \in X \mid d(x, Y) < r\} \quad \bar{N}_r(Y) := \{x \in X \mid d(x, Y) \leq r\}$$

In the latter two we mean $d(x, Y) := \inf_{y \in Y} d(x, y)$.

1.2. Paths, length spaces, and geodesics

A *path* is a continuous map $\gamma: I \rightarrow X$, where $I \subseteq \mathbb{R}$ is an interval.

The *length* of a path $\gamma: [a, b] \rightarrow X$ is:

$$|\gamma| := \sup \left\{ \sum_{i=0}^{n-1} d(\gamma(t_i), \gamma(t_{i+1})) \mid a = t_0 < t_1 < \dots < t_n = b \right\}$$

A path is *rectifiable* if its length is finite, and *locally rectifiable* if its restriction to every compact subinterval of its domain has finite length.

LEMMA 1.2.1 (Properties of length). *For a locally rectifiable path $\gamma: I \rightarrow X$ and $r \leq s \leq t$ in I :*

- $d(\gamma(s), \gamma(t)) \leq |\gamma|_{[s, t]}$
- $|\gamma|_{[r, s]} + |\gamma|_{[s, t]} = |\gamma|_{[r, t]}$
- *There is a unit-speed reparametrization: For each $[s, t] \subseteq I$ there is a continuous, weakly monotonic function $\phi: [s, t] \rightarrow [0, |\gamma|_{[s, t]}]$ and a unique path $\tilde{\gamma}: [0, |\gamma|_{[s, t]}] \rightarrow X$ such that $\tilde{\gamma} \circ \phi = \gamma$ and $|\tilde{\gamma}|_{[0, u]} = u$ for all $u \in [0, |\gamma|_{[s, t]}]$.*

DEFINITION 1.2.2. A metric space (X, d) is a *length space* if, for all $x, y \in X$ and every $\varepsilon > 0$, there exists a rectifiable path γ from x to y of length less than $d(x, y) + \varepsilon$.

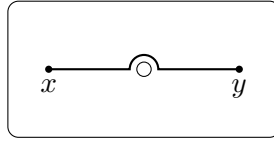
A *geodesic* is a locally rectifiable path $\gamma: I \rightarrow X$ such that $\forall s, t \in I$:

$$|\gamma|_{[s,t]} = d(\gamma(s), \gamma(t))$$

DEFINITION 1.2.3. A metric space (X, d) is a *geodesic space* if every pair $x, y \in X$ can be joined by a geodesic.

EXAMPLE 1.2.4. A space that is not path-connected is not a length space.

EXAMPLE 1.2.5 (Punctured plane). Consider the plane minus a point, $\mathbb{R}^2 \setminus \{\vec{0}\}$, with the restriction of the Euclidean metric. For points x, y whose Euclidean segment passes through the puncture, any path between them in the punctured plane must detour around the puncture, which costs some additional length. We can make the detour as small as desired, but not zero. This space is a length space but not a geodesic space.



1.3. Graphs and trees

A *graph* Γ consists of a vertex set $V\Gamma$ and an edge set $E\Gamma$. Multiple edges and loops are allowed. Regard every edge as having length 1; the induced path-length distance is a metric precisely when Γ is connected, and then it is geodesic.

An *edge path* has the form:

$$v_0 \xrightarrow{e_1} v_1 \xrightarrow{e_2} \cdots \xrightarrow{e_n} v_n$$

A *cycle* is a closed edge path ($v_0 = v_n$) without repeated edges or other repeated vertices.

DEFINITION 1.3.1. A *tree* is a nonempty connected graph without cycles.

LEMMA 1.3.2. A finite tree that is not a single vertex has at least two leaves, that is, vertices of valence 1.

PROOF. Choose a longest embedded edge path $v_0 \xrightarrow{e_1} \cdots \xrightarrow{e_n} v_n$. We claim that v_0 and v_n are leaves. If, say, v_0 had another incident edge f with opposite endpoint w , then either $w \notin \{v_0, \dots, v_n\}$, giving a longer embedded path, or $w \in \{v_0, \dots, v_n\}$, giving a cycle. $\square$

COROLLARY 1.3.3. If Γ is a tree, then every nonconstant closed edge path backtracks at least twice (counted cyclically, so allowing the possibility that one of the backtracks is at the initial/terminal vertex).

PROOF. The image is a connected finite subtree, hence has at least two leaves. At each leaf the path must backtrack. $\square$

COROLLARY 1.3.4. If Γ is a tree and $x \neq y$ are vertices, there is a unique embedded edge path from x to y .

PROOF. Suppose α, β are embedded paths from x to y . Let $\bar{\beta}$ be the edge path β traversed in the opposite direction. The concatenation $\alpha + \bar{\beta}$ is a closed edge path, so it has at least two backtracks. Since α and β have no internal backtracking, these must occur at the endpoints x, y . Removing the common initial and terminal edges and inducting on $d(x, y)$ gives $\alpha = \beta$. $\square$

LEMMA 1.3.5. For a connected graph Γ , the following are equivalent:

- (i) Γ is a tree.
- (ii) Γ has no embedded circle $\mathbb{S}^1$.
- (iii) Γ has no loops and every pair of vertices is joined by a unique embedded edge path.

- (iv) Every pair of points is joined by a unique geodesic.
- (v) For each point x , either $\Gamma = \{x\}$, or x is a leaf, or x is a cut point of Γ .

For points x, y of a tree, write $[x, y]$ for the unique geodesic joining them.

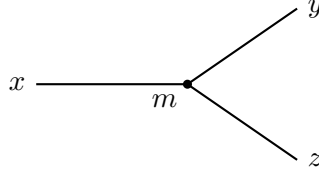
LEMMA 1.3.6. *If Γ is a tree, $x, y \in \Gamma$, and γ is any path from x to y , then the image of γ contains $[x, y]$.*

PROOF. The image of γ contains x and y . If $x = y$ we are done. Otherwise, every interior point z of $[x, y]$ is a cut point of Γ . If z is an interior point of an edge then locally it only has two complementary components, so these identify the two global complementary components. Thus, x and y are separated by z , meaning that every path from x to y passes through z . Since the image of γ is closed and contains the interior points of every edge on $[x, y]$, it also contains the vertices. $\square$

LEMMA 1.3.7 (Tripod lemma). *For any x, y, z in a tree there exists a unique point m such that:*

$$[x, y] \cap [y, z] \cap [x, z] = \{m\}$$

PROOF. Let m be the last common point of $[x, y]$ and $[x, z]$. The two geodesic segments $[y, m]$ and $[m, z]$ concatenate without backtracking, since a backtrack would contradict the choice of m . Thus their concatenation is $[y, z]$, and the triple intersection is exactly $\{m\}$. $\square$



$\square$

THEOREM 1.3.8. *Geodesic triangles in a tree are 0-thin: if $x_0, x_1, x_2 \in \Gamma$ and γ_i is a geodesic from x_i to x_{i+1} (indices modulo 3), then for all i :*

$$\gamma_i \subseteq \gamma_{i+1} \cup \gamma_{i+2} = \bar{N}_0(\gamma_{i+1} \cup \gamma_{i+2})$$

PROOF. Every triangle is a tripod by the preceding lemma. If m is its central point, then for each i :

$$[x_i, x_{i+1}] = [x_i, m] \cup [m, x_{i+1}] \subseteq [x_i, x_{i+2}] \cup [x_{i+2}, x_{i+1}] \quad \square$$

1.4. The Gromov product and the boundary of a tree

DEFINITION 1.4.1. The *Gromov product* of x and y based at z is:

$$(x \mid y)_z := \frac{1}{2} (d(x, z) + d(y, z) - d(x, y))$$

If T is a tree and m is the center of the tripod with vertices x, y, z , then

$$(x \mid y)_z = d(m, z) = d(z, [x, y])$$

Let T be a possibly unbounded tree. A (geodesic) *ray* is an embedded edge path that is infinite in one direction:

$$v_0 \xrightarrow{e_1} v_1 \xrightarrow{e_2} v_2 \xrightarrow{\quad} \cdots$$

DEFINITION 1.4.2. Two rays are *asymptotic* if their intersection contains a ray, that is, if they eventually agree. The *boundary* ∂T is the set of asymptotic equivalence classes of rays.

Fix a base vertex $o \in T$.

LEMMA 1.4.3. *For each $\xi \in \partial T$, there is exactly one ray starting at o and representing ξ .*

PROOF. Take a ray $\alpha \in \xi$ and let $\alpha(t_0)$ be the last vertex of α on the segment $[o, \alpha(0)]$. Then $[o, \alpha(t_0)] + \alpha|_{[t_0, \infty)}$ is a ray representing ξ . Uniqueness follows from uniqueness of embedded paths in a tree. $\square$

Extending the Gromov product to ∂T . For $\xi, \eta \in \partial T$, define:

$(\xi | \eta)_o := \text{length of the common initial segment of the rays } \alpha_\xi \in \xi \text{ and } \alpha_\eta \in \eta \text{ based at } o$

LEMMA 1.4.4. *If $\xi \neq \eta \in \partial T$, there is a unique bi-infinite geodesic (ξ, η) such that each of its two rays from any vertex represents one of ξ or η . Moreover,*

$$(\xi | \eta)_o = d(o, (\xi, \eta))$$

THEOREM 1.4.5. *The following defines a metric on ∂T :*

$$d(\xi, \eta) := 2^{-(\xi | \eta)_o}$$

PROOF. Symmetry and nonnegativity are immediate. Also,

$$d(\xi, \eta) = 0 \iff (\xi | \eta)_o = \infty \iff \alpha_\xi = \alpha_\eta \iff \xi = \eta$$

For the triangle inequality, the common initial segment of α_ξ and α_ζ has length at least the minimum of the common initial segment lengths for $(\alpha_\xi, \alpha_\eta)$ and $(\alpha_\eta, \alpha_\zeta)$. Thus:

$$(\xi | \zeta)_o \geq \min\{(\xi | \eta)_o, (\eta | \zeta)_o\}$$

This gives the stronger ultrametric inequality:

$$d(\xi, \zeta) \leq \max\{d(\xi, \eta), d(\eta, \zeta)\} \quad \square$$

1.5. Cantor boundaries and compactification

THEOREM 1.5.1. *Let T be a tree in which every vertex has finite valence at least 3. Then ∂T is the Cantor set.*

PROOF. We use the fact that the Cantor set is characterized as the unique nonempty compact metrizable totally disconnected perfect topological space.

A tree without leaves is unbounded, so its boundary is nonempty. We use the metric topology given by Theorem 1.4.5, so metrizability is automatic.

Compactness. Since ∂T is metrizable, compactness is equivalent to sequential compactness. Let (ξ_i) be a sequence in ∂T , and let α_{ξ_i} be the corresponding rays based at o . Only finitely many edges are incident to o , so we can pass to a subsequence whose rays have the same first edge. The opposite vertex to o also has finite valence, so we can pass to a further subsequence with the same first two edges, and continue. A diagonal subsequence converges to the ray α whose first n edges are the stabilized initial n -edge segments. Thus $\xi_{i_k} \rightarrow [\alpha]$.

No isolated points (perfectness). Given $\xi \in \partial T$ and $n \geq 1$, take a ray α representing ξ and branch away from it after n edges. Since every vertex has valence at least 3, this produces a distinct ray α_n agreeing with α on the first n edges. Hence $[\alpha_n] \rightarrow \xi$.

Totally disconnected. For $\xi \neq \eta$ choose n so that $2^{-n} < d(\xi, \eta)$, and set

$$C_n(\xi) := \{\zeta \in \partial T \mid (\xi | \zeta)_o \geq n\}$$

This is the closed ball $\overline{B}_{2^{-n}}(\xi)$, and it is also open. Its complement is open because sufficiently close rays must share the same initial n edges. Thus ξ and η are separated by a clopen set and cannot lie in the same connected component. $\square$

THEOREM 1.5.2 (Compactification). *Let T be a tree of finite valence at least 3. There is a topology on $\bar{T} := T \cup \partial T$ with the following properties:*

- (i) $\bar{T}$ is compact.
- (ii) Its restrictions to T and ∂T agree with their respective metric topologies.
- (iii) T is open and dense in $\bar{T}$.

PROOF. For $x \in T$ use the usual metric balls as a neighborhood basis. For $\xi \in \partial T$, choose a ray α_ξ based at o and take as a neighborhood basis the sets:

$$\begin{aligned} C_n(\xi) &:= \{x \in T \mid [o, x] \text{ and } \alpha_\xi \text{ agree for length } \geq n\} \\ &\cup \{\eta \in \partial T \mid (\xi | \eta)_o \geq n\} \end{aligned}$$

Each boundary point is represented by a ray, and the vertices of the ray approach the boundary point in this topology, so T is dense. For every point $x \in T$ and every positive r the ball $B_r(x)$ is an open set containing x and contained in T , so T is open.

Compactness is proved by the same diagonal argument as for ∂T . (It turns out that $\bar{T}$ is metrizable, but here it is enough to observe that $\bar{T}$ is second countable, so compact is equivalent to sequentially compact.) Every sequence in ∂T has a convergent subsequence. For a sequence (x_i) in T , either it is bounded, in which case properness of the locally finite tree gives a convergent subsequence, or (after passing to a subsequence) $d(o, x_i) \rightarrow \infty$. In the latter case, apply the diagonal argument to the finite segments $[o, x_i]$ to obtain a limiting ray. $\square$

1.6. Isometries of regular trees

Let T_r be a regular tree of valence r , where $3 \leq r < \infty$. An *isometry* $\phi: T_r \rightarrow T_r$ preserves distances:

$$d(x, y) = d(\phi(x), \phi(y)) \quad (x, y \in T_r)$$

Notice that vertices have valence at least 3, vertices are distinguishable from interior points of edges. Thus, a map $T_r \rightarrow T_r$ is an isometry if and only if it is an isomorphism of the tree structure, meaning a bijection of the vertices that preserves adjacency.

An isometry is one of the following types:

Elliptic: It fixes a point of T_r . We will show in this case it fixes a vertex or the midpoint of an edge.

Hyperbolic/loxodromic: it has no fixed point in T_r ; it fixes exactly two boundary points ξ^-, ξ^+ , translates along the axis (ξ^-, ξ^+) , and has north-south dynamics on ∂T_r : for every pair of disjoint neighborhoods $U^- \ni \xi^-$, $U^+ \ni \xi^+$, there exists N such that $\forall n \geq N$:

$$\phi^n(\partial T_r \setminus U^-) \subseteq U^+$$

Also, the axis (ξ^-, ξ^+) is equal to the minimal displacement set of ϕ .

PROOF OF THE CLASSIFICATION. If ϕ inverts an edge then it fixes the midpoint of that edge.

Suppose ϕ fixes no vertex and inverts no edge. Suppose u is an interior point of an edge $w -_e x$. Then $\phi(u) \notin e$, so it is an interior point of some different edge $y - z$. Consider the geodesic from u to $\phi(u)$. It goes through either w or x , and it goes through either y or z . Up to relabelling, suppose it goes through x and y . Then $d(x, y) < d(u, \phi(u))$. Now there are two cases: we have $\phi(\{w, x\}) = \{y, z\}$, so either $\phi(x) = y$ or $\phi(x) = z$. In the former case $d(x, \phi(x)) < d(u, \phi(u))$. In the latter case $d(x, \phi(x)) = d(u, \phi(u))$. Thus, for all u there exists a vertex x such that $d(x, \phi(x)) \leq d(u, \phi(u))$.

Since vertices move by distances that are nonnegative integers, we have shown that the following quantity is a minimum, and it is either 0 and realized at the midpoint of an edge, or it is realized at a vertex:

$$\tau(\phi) := \inf_{u \in T_r} d(u, \phi(u))$$

The elliptic case is $\tau(\phi) = 0$, which is realized at either the midpoint of an edge or at a vertex.

For the loxodromic case, suppose $\tau(\phi) > 0$ and choose $v \in VT_r$ realizing the positive minimum. We claim that the segments $[v, \phi(v)]$ and $[\phi(v), \phi^2(v)]$ do not overlap in a nontrivial segment. Otherwise, let x be the point of $[v, \phi(v)]$ at distance $1/2$ from $\phi(v)$. Then x is also the point of $[\phi(v), \phi^2(v)]$ at distance $1/2$ from $\phi(v)$, and $\phi(x)$ is the point of $[\phi(v), \phi^2(v)]$ at distance $1/2$ from $\phi^2(v)$, so:

$$d(x, \phi(x)) = d(\phi(v), \phi^2(v)) - 1 = d(v, \phi(v)) - 1$$

This contradicts minimality of $d(v, \phi(v))$.

Consequently the segments

$$[\phi^n(v), \phi^{n+1}(v)] \quad (n \in \mathbb{Z})$$

concatenate to form a bi-infinite geodesic. The isometry ϕ translates along this axis, and its two endpoints are

$$\xi^+ := \lim_{n \rightarrow \infty} \phi^n(v) \quad \xi^- := \lim_{n \rightarrow -\infty} \phi^n(v)$$

If η is different from these endpoints, the geodesic ray from σ to $\phi^n(\eta)$ overlaps longer and longer with the ray toward ξ^+ , uniformly for η outside any fixed neighborhood of ξ^- . This is the asserted north-south dynamics.

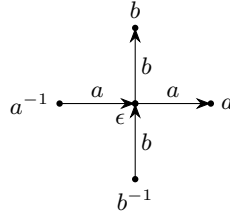
Finally, the axis (ξ^-, ξ^+) is the entire minimal displacement set, as follows. If $u \notin (\xi^-, \xi^+)$ then there is closest point v of (ξ^-, ξ^+) to u . The point v is a vertex, and $[u, v] \cap (\xi^-, \xi^+) = \{v\}$. Apply ϕ : $\phi(v)$ is the closest point of (ξ^-, ξ^+) to $\phi(u)$, $\phi([u, v]) = [\phi(u), \phi(v)]$, and $[\phi(u), \phi(v)] \cap (\xi^-, \xi^+) = \{\phi(v)\}$. Then we claim that v and $\phi(v)$ separate u and $\phi(u)$, so $d(u, \phi(u)) > d(v, \phi(v))$. $\square$

1.7. The free group and its Cayley tree

The group $F_2 = \langle a, b \mid \rangle$ is the free group on two generators. Its elements are uniquely represented by freely reduced words in the letters a, a^{-1}, b, b^{-1} . The identity is the empty word, denoted ϵ . Multiplication is concatenation followed by free reduction. For example,

$$(aba^{-1})(ab^2) = ab^3$$

The group F_2 acts isometrically on the regular tree T_4 . The action is free and transitive on the vertices, has no edge inversions, and has two orbits of unoriented edges. We establish these facts by constructing a labelling of the tree. Label one vertex by ϵ . It has 4 incident edges. We declare two of them to be outgoing and label them a and b , and declare the other two to be incoming, and label them a^{-1} and b^{-1} . By convention, the label of an edge read against its direction is the inverse of the given label, so from the point of view of ϵ the four outgoing edges are labelled a, b, a^{-1} and b^{-1} . Label the corresponding vertices a, b, a^{-1} and b^{-1} .



Now induct: take a closest vertex to ϵ that is labelled, say by reduced word w , but for which not all of its incident edges are labelled. By construction, exactly one of its incident edges has been directed and labelled, the edge leading to ϵ . Assign directions and labels to the remaining three edges in such a way that vertex w has one incoming and one outgoing edge labelled by each of a and b . Label the adjacent vertices of w as follows. For the outgoing edge labelled a , label the next vertex by the free reduction of wa . For the incoming edge labelled a , label the next vertex by the free reduction of wa^{-1} . Proceed similarly for b . Observe that there is exactly one of these four labels that require free reduction: suppose, for example, that $w = ua$. Then appending a, b , or b^{-1} results in a freely reduced word; only a^{-1} requires a free reduction. Moreover, we arrived at the vertex labelled w from the vertex labelled u by following the outgoing edge at u labelled a , so $wa^{-1} = uaa^{-1} \mapsto u$. Thus, the labels assigned in the induction step are consistent with labels from previous steps.

This process limits to a labelling of the tree in which every vertex is labelled by a freely reduced word of length equal to its distance from ϵ , and every vertex has one incoming and one outgoing edge with each label a and b . In such a labelled tree, an edge path has a label, defined by concatenating the labels of its edges, inverting the label if the edge is traversed against its assigned direction. The label of an embedded edge path is a freely reduced word. Moreover, given a freely reduced word and a vertex, there is a unique edge path starting at that vertex whose label is the given reduced word. It follows that the labelling gives a bijection between vertices of T_4 and elements of F_2 , since elements of F_2 are in bijection with freely reduced words,

which are in bijection with embedded edge paths with starting vertex ϵ , thus with vertices of T_4 .

Now we claim that left multiplication of F_2 on the labels of the vertices induces an isometry of T_4 . This is because edges are of the form $x \xrightarrow{a} xa$, so $w \in F_2$ takes the end vertices of that edge to wx and wxa , which differ by right multiplication of a , so there is an edge $wx \xrightarrow{a} wxa$. Thus, F_2 acts on T_4 , freely and transitively on vertices, and preserving the directions and labels of the edges.

Suppose $\epsilon \neq w = uvu^{-1}$ is written as a freely reduced product with v cyclically reduced. Thus there is no cancellation within uvu^{-1} , and the first and last letters of v are not inverses. Multiplication by w has fixed points in ∂T_4 :

$$w^{+\infty} = uv^{+\infty} := \lim_{n \rightarrow \infty} uv^n \quad w^{-\infty} = uv^{-\infty} := \lim_{n \rightarrow \infty} uv^{-n}$$

These are the two ends of the axis of w .

1.8. Centralizers in the free group

THEOREM 1.8.1. *For $w \in F_2 \setminus \{1\}$, the centralizer*

$$Z_{F_2}(w) := \{x \in F_2 \mid xwx^{-1} = w\}$$

is isomorphic to $\mathbb{Z}$.

PROOF. Let $x \in Z_{F_2}(w)$. The isometry w translates along the bi-infinite axis:

$$A_w = (w^{-\infty}, w^{+\infty})$$

For any $v \in A_w$, we claim that $xv \in A_w$, because A_w is the entire minimal displacement set of the w -action, and:

$$\tau(w) = d(v, wv) = d(xv, xwv) = d(xv, wxv)$$

Therefore $x(A_w) = A_w$. It cannot interchange the two endpoints of A_w , because then it would fix an interior point of the axis, whereas the F_2 action has no nontrivial elliptic elements. Hence x acts on A_w by translation.

Restriction to the axis gives a homomorphism

$$Z_{F_2}(w) \longrightarrow \text{Trans}(A_w) \cong \mathbb{Z}$$

Its kernel is trivial because the action of F_2 on vertices of T_4 is free. The image is a subgroup of $\mathbb{Z}$ and is nontrivial because it contains w . Hence the centralizer is isomorphic to $\mathbb{Z}$. $\square$

1.9. Ping-pong on the boundary

THEOREM 1.9.1 (Ping-pong). *Let $f, g \in F_2 \setminus \{1\}$ satisfy*

$$|\{f^{+\infty}, f^{-\infty}, g^{+\infty}, g^{-\infty}\}| = 4$$

Then, for some $n \geq 1$, the subgroup generated by f^n and g^n is isomorphic to F_2 .

PROOF. Choose pairwise disjoint neighborhoods

$$A^+ \ni f^{+\infty}, \quad A^- \ni f^{-\infty}, \quad B^+ \ni g^{+\infty}, \quad B^- \ni g^{-\infty}$$

which do not together cover all of ∂T_4 . By north-south dynamics, choose n large enough that

$$\begin{aligned} f^n(\partial T_4 \setminus A^-) &\subseteq A^+ \\ f^{-n}(\partial T_4 \setminus A^+) &\subseteq A^- \\ g^n(\partial T_4 \setminus B^-) &\subseteq B^+ \\ g^{-n}(\partial T_4 \setminus B^+) &\subseteq B^- \end{aligned}$$

Set $\alpha = f^n$ and $\beta = g^n$. We wish to prove that every nonempty freely reduced word w in $\alpha^{\pm 1}, \beta^{\pm 1}$ is nontrivial. Choose $\xi \in \partial T_4$ outside all four sets A^+, A^-, B^+, B^- . It suffices to show that $w\xi \neq \xi$.

The point ξ is the “ping-pong ball”, and the game is that the letters of w hit the ball into their designated regions: If, for example, the last letter of w is α , so $w = w'\alpha$, then the first player is α , and, since $\xi \notin A^-$, α hits ξ into A^+ , ie, $\alpha\xi \in A^+$. The next player is the last letter of w' . Since w is freely reduced, the next player is not α^{-1} , it is one of α , β , or β^{-1} . Since the ball currently lies in A^+ , it is not in A^- , B^- , or B^+ , so player 2 hits the ball into A^+ , B^+ , or B^- , respectively. Continue, inducting on the length of w , to see that from the first hit onwards, the ball is always in one of the sets A^+ , A^- , B^+ , or B^- , so it never returns to the starting point ξ . Thus, $w\xi \neq \xi$, so $w \neq 1$. $\square$

1.10. Exercises

EXERCISE 1.1. Prove Lemma 1.2.1.

EXERCISE 1.2. Prove Lemma 1.3.5.

EXERCISE 1.3. Prove Lemma 1.4.4.

EXERCISE 1.4. **Gromov products and the four-point condition.**

Let T be a tree and $o \in T$.

(a) Show that for every $x, y, z \in T$:

$$(x \mid z)_o \geq \min\{(x \mid y)_o, (y \mid z)_o\}$$

(b) Show that for any four points $x, y, z, w \in T$, the two largest of the following quantities are equal:

$$d(x, y) + d(z, w)$$

$$d(x, z) + d(y, w)$$

$$d(x, w) + d(y, z)$$

(c) Interpret this exercise as a statement about geodesic quadrilaterals in a tree.

EXERCISE 1.5. **Nearest-point projections.**

Let C be a nonempty connected subgraph of a tree T .

(a) Show C is a tree.

(b) Show that every $x \in T$ has a unique closest point $\pi_C(x) \in C$.

(c) Show that for every $y \in C$, the geodesic $[x, y]$ passes through $\pi_C(x)$.

(d) Show that $\pi_C: T \rightarrow C$ is 1-Lipschitz. A map ϕ is L -Lipschitz if $d(\phi(x), \phi(y)) \leq Ld(x, y)$ for all x and y .

EXERCISE 1.6. **The Morse lemma for trees.**

A (not necessarily continuous!) map $\gamma: I \rightarrow X$ is a (λ, ε) -quasigeodesic if $\lambda \geq 1$, $\varepsilon \geq 0$, and for all s, t in the interval I :

$$\frac{1}{\lambda}|s - t| - \varepsilon \leq d(\gamma(s), \gamma(t)) \leq \lambda|s - t| + \varepsilon$$

Suppose γ is a continuous (λ, ε) -quasigeodesic from x to y in a tree T .

(a) Use Lemma 1.3.6 to show that $[x, y]$ is contained in the image of γ .

(b) For $z \in \gamma([a, b]) \setminus [x, y]$, let p be its closest point on $[x, y]$. Show that there exist $s < t < u$ such that $\gamma(s) = \gamma(u) = p$ and $\gamma(t) = z$.

(c) Deduce that:

$$d(z, [x, y]) \leq \frac{\lambda^2 \varepsilon}{2} + \varepsilon$$

(d) Show that every $(\lambda, 0)$ -quasigeodesic in a tree has the same image as the geodesic between its endpoints.

EXERCISE 1.7. **Changing the basepoint.**

Let T be a tree and let $o, o' \in VT$.

(a) Show that for $\xi, \eta \in \partial T$:

$$|(\xi \mid \eta)_o - (\xi \mid \eta)_{o'}| \leq d(o, o')$$

(b) Deduce that the two metrics given by Theorem 1.4.5 satisfy:

$$2^{-d(o,o')}d_o(\xi,\eta) \leq d_{o'}(\xi,\eta) \leq 2^{d(o,o')}d_o(\xi,\eta)$$

(c) Conclude that the topology of ∂T is independent of the choice of basepoint.

EXERCISE 1.8. Translation lengths.

Let g be a hyperbolic isometry of T_r , with axis A_g .

(a) Prove that for every $n \in \mathbb{Z}$:

$$\tau(g^n) = |n|\tau(g)$$

(b) Prove that for every $x \in T_r$ and every positive integer n :

$$d(x, g^n x) = n\tau(g) + 2d(x, A_g)$$

(c) Deduce that:

$$\lim_{n \rightarrow \infty} \frac{d(x, g^n x)}{n} = \tau(g)$$

EXERCISE 1.9. North–south dynamics.

Give a detailed proof of the north–south dynamics assertion from the classification of isometries of T_r .

EXERCISE 1.10. Common boundary fixed points.

Let $f, g \in F_2 \setminus \{1\}$.

- (a) Suppose that f and g have a common fixed point in ∂T_4 . Show that their axes have a common infinite ray.
- (b) Show that the commutator $[f, g]$ fixes a vertex of T_4 .
- (c) Deduce that f and g commute.
- (d) Conclude that f and g belong to a common cyclic subgroup of F_2 , and that they have the same two fixed points in ∂T_4 .
- (e) Deduce from Theorem 1.9.1 that if f and g do not belong to a common cyclic subgroup, then for sufficiently large n :

$$\langle f^n, g^n \rangle \cong F_2$$

LECTURE 2

The hyperbolic plane and surface groups

2.1. The Poincaré disk and its metric

In the first lecture we had a metric space definition of the length of a continuous path. Recall from calculus that the length of a C^1 curve $\gamma: [a, b] \rightarrow \mathbb{R}^n$ is:

$$\int_a^b |\gamma'(t)| dt$$

The Euclidean length of $\gamma'(t)$ is measured using the usual inner product. We can instead choose a different inner product for tangent vectors at each point.

DEFINITION 2.1.1 (Poincaré disk). Let $\mathbb{D} := \{z \in \mathbb{C} \mid |z| < 1\}$. At $z \in \mathbb{D}$ define an inner product on tangent vectors $v, w \in T_z \mathbb{D} \cong \mathbb{C}$ by:

$$\langle v, w \rangle_z := \frac{4 \operatorname{Re}(v\bar{w})}{(1 - |z|^2)^2} = \frac{4 \langle v, w \rangle_{\mathbb{E}^2}}{(1 - |z|^2)^2}$$

Thus the length of a C^1 curve $\gamma: [a, b] \rightarrow \mathbb{D}$ is:

$$L_{\mathbb{H}}(\gamma) := \int_a^b \frac{2|\gamma'(t)|}{1 - |\gamma(t)|^2} dt$$

Define a metric on $\mathbb{D}$ by:

$$d_{\mathbb{H}}(z, w) := \inf\{L_{\mathbb{H}}(\gamma) \mid \gamma \text{ is piecewise } C^1 \text{ curve joining } z \text{ to } w\}$$

The resulting length space is the *hyperbolic plane* $\mathbb{H}^2$. This specific description of the hyperbolic plane is the *Poincaré disk model*.

The factor 4 normalizes the sectional curvature to -1 . For us, the important thing is that the curvature is negative, not what its magnitude is. For comparison, the Euclidean plane has sectional curvature 0, and the unit sphere with its usual metric has sectional curvature $+1$.

Notice that at every point z the inner product on $T_z \mathbb{D}$ is a rescaling of the Euclidean inner product. The lengths of the vectors are different, but angles between them are the same as in the Euclidean sense. In a small neighborhood around z all tangent planes are scaled by close factors, so on small neighborhoods the metric behaves much like the Euclidean metric.

EXAMPLE 2.1.2. For $a < 1$, let $\gamma: [0, a] \rightarrow \mathbb{H}^2 : t \mapsto t$. We have $\gamma'(t) = 1$ for all t . In Euclidean terms γ has constant unit speed.

$$\begin{aligned} L_{\mathbb{H}^2}(\gamma) &= \int_0^a |\gamma'(t)|_{T_{\gamma(t)} \mathbb{H}^2} dt \\ &= \int_0^a \frac{2|\gamma'(t)|_{\mathbb{E}^2}}{1 - |\gamma(t)|^2} dt \\ &= 2 \int_0^a \frac{1}{1 - t^2} dt \\ &= \log \left. \frac{1+t}{1-t} \right|_0^a \\ &= \log \frac{1+a}{1-a} \end{aligned}$$

Notice as a gets close to 1, this curve becomes arbitrarily long.

We cannot determine the distance from 0 to a from this example alone. We have only computed the length of one curve. Potentially there could be some shorter curve with the same endpoints. It turns out that that is not the case.

PROPOSITION 2.1.3. *Projection to any line through 0 in $\mathbb{H}^2$ is length non-increasing.*

PROOF. Let π be orthogonal projection to a line L . Choose a unit vector in the direction of L and extend it to an orthonormal basis of $\mathbb{R}^2$. Let γ be any differentiable curve. The tangent vector to $\pi \circ \gamma$ at t in the new coordinates is the same as $\gamma'(t)$, but with the second coordinate set to 0, so its Euclidean length does not increase. Furthermore, $\pi \circ \gamma(t)$ is no farther from the origin than $\gamma(t)$, so $|\pi \circ \gamma|_{\mathbb{H}^2} \leq |\gamma|_{\mathbb{H}^2}$. $\square$

PROPOSITION 2.1.4 (Radial distance and geodesics). *For $z \in \mathbb{D}$:*

$$d_{\mathbb{H}}(0, z) = 2 \int_0^{|z|} \frac{dr}{1-r^2} = \log \frac{1+|z|}{1-|z|}$$

The radial segment from 0 to z is the unique geodesic between its endpoints, up to reparametrization.

PROOF. The displayed integral gives the length of the radial segment. Every curve from 0 to z projects onto a curve in the diameter through z with no greater length. This projected curve must traverse the radial segment and hence has length at least the displayed quantity. Equality is possible only if the original curve stays on that diameter and has no backtracking. $\square$

In particular, points approaching the boundary circle are infinitely far from the origin:

$$d_{\mathbb{H}}(0, z) \rightarrow \infty \quad \text{as } |z| \rightarrow 1$$

Recall:

$$\sinh(x) := \frac{e^x - e^{-x}}{2} \quad \cosh(x) := \frac{e^x + e^{-x}}{2} \quad \tanh(x) := \frac{\sinh(x)}{\cosh(x)} = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

LEMMA 2.1.5. *For $z \in \mathbb{D}$:*

$$d_{\mathbb{H}^2}(0, z) = 2 \operatorname{artanh}(|z|)$$

Equivalently:

$$d_{\mathbb{H}^2}(0, z) = 2 \operatorname{arsinh} \left(\frac{|z|}{\sqrt{1-|z|^2}} \right)$$

PROOF. A direct computation using Proposition 2.1.4 yields the first. The second follows from the identity $\sinh(\operatorname{artanh}(r)) = \frac{r}{\sqrt{1-r^2}}$. $\square$

2.2. The isometry group

We take the following theorem from Riemannian geometry as a black box:

THEOREM 2.2.1 (Myers–Steenrod). *A distance-preserving bijection between connected Riemannian manifolds is smooth. Moreover, an isometry is uniquely determined by its value and its derivative at one point.*

Rotations $z \mapsto e^{i\theta}z$ and complex conjugation $z \mapsto \bar{z}$ clearly preserve the hyperbolic metric: they preserve Euclidean lengths and distances from 0. But all of those maps fix 0.

PROPOSITION 2.2.2. *Consider:*

$$\operatorname{PSU}(1, 1) := \left\{ \begin{bmatrix} a & b \\ \bar{b} & \bar{a} \end{bmatrix} \mid \begin{pmatrix} a & b \\ \bar{b} & \bar{a} \end{pmatrix} \in \mathcal{M}_{2 \times 2}(\mathbb{C}), |a|^2 - |b|^2 = 1 \right\}$$

- (1) $\operatorname{PSU}(1, 1)$ acts on $\mathbb{H}^2$.
- (2) $\operatorname{PSU}(1, 1)$ contains $\operatorname{SO}(2) < \operatorname{Isom}^+(\mathbb{H}^2)$.
- (3) $\operatorname{PSU}(1, 1)$ contains a 1-parameter family that acts by hyperbolic translation on the real axis.
- (4) $\operatorname{PSU}(1, 1)$ acts transitively on the set of unit tangent vectors to $\mathbb{H}^2$.

(5) $\text{Isom}^+(\mathbb{H}^2)$ acts simply transitively on the set of unit tangent vectors to $\mathbb{H}^2$.

Consequently, $\text{Isom}^+(\mathbb{H}^2) \cong \text{PSU}(1, 1)$.

PROOF. Let $\text{SU}(1, 1) := \left\{ \begin{pmatrix} a & b \\ \bar{b} & \bar{a} \end{pmatrix} \in \mathcal{M}_{2 \times 2}(\mathbb{C}) \mid |a|^2 - |b|^2 = 1 \right\}$ act on $\mathbb{C} \cup \{\infty\}$ by Möbius transformations, so that:

$$\begin{pmatrix} a & b \\ \bar{b} & \bar{a} \end{pmatrix} : z \mapsto \frac{az + b}{\bar{b}z + \bar{a}}$$

Composition of Möbius transforms is compatible with matrix multiplication, so such a map is a bijection of $\mathbb{C} \cup \{\infty\}$. Observe that the center of $\text{SU}(1, 1)$ consists of plus and minus the identity matrix, and this is also the kernel of the action. $\text{PSU}(1, 1) := \text{SU}(1, 1) / \pm \text{Id}$.

For $c \in \mathbb{D}$, let $F_c(z) := \frac{z+c}{\bar{c}z+1}$. Note that if $|a|^2 - |b|^2 = 1$ then $a \neq 0$ and $|b/a|^2 = \frac{|b|^2}{1+|b|^2} < 1$. Then:

$$\frac{az + b}{\bar{b}z + \bar{a}} = \frac{a}{\bar{a}} \frac{z + b/a}{(\bar{b}/\bar{a})z + 1} = e^{i\theta} F_c(z) \quad \text{for} \quad \theta := \arg\left((a/|a|)^2\right) \text{ and } c := b/a$$

$F_c(z)$ sends 0 to c and has inverse $F_{-c}(z)$.

$$1 - |F_c(z)|^2 = \frac{(1 - |c|^2)(1 - |z|^2)}{|1 + \bar{c}z|^2} > 0$$

F_c is holomorphic on $\mathbb{D}$, so the derivative $D_z F_c: T_z \mathbb{D} \rightarrow T_{F_c(z)} \mathbb{D}$ is given by complex multiplication by $F'_c(z)$.

$$|F'_c(z)| = \left| \frac{1(\bar{c}z + 1) - \bar{c}(z + c)}{(\bar{c}z + 1)^2} \right| = \frac{1 - |c|^2}{|\bar{c}z + 1|^2}$$

Consequently, for $v \in T_z \mathbb{D}$:

$$\sqrt{\langle D_z F_c(v), D_z F_c(v) \rangle_{F_c(z)}} = \frac{2|F'_c(z)v|}{1 - |F_c(z)|^2} = \frac{2|v|}{1 - |z|^2} = \sqrt{\langle v, v \rangle_z}$$

So F_c preserves the hyperbolic length of tangent vectors, so it preserves hyperbolic lengths of curves, hence the hyperbolic metric.

A copy of $\text{SO}(2)$ fixing 0 is obtained by taking $b = c = 0$.

Translations along the real axis are obtained by taking $\theta = 0$ and real c , which is equivalent to a and b being real. For translation by hyperbolic distance t , take $c := \tanh(t/2)$.

For any $z \in \mathbb{D}$ and $v \in T_z \mathbb{D}$, the isometry F_{-z} takes z to 0 and its derivative takes v to a vector in $F'_{-z}v \in T_0 \mathbb{D}$ of the same hyperbolic norm as v . If $\theta := -\arg(F'_{-z}v)$ then $e^{i\theta} F_{-z}$ is an isometry taking z to 0 and taking v to the unique vector in $T_0 \mathbb{D}$ pointing in the positive real direction and with hyperbolic norm equal to that of v .

Finally, we claim that if $\phi \in \text{Isom}^+(\mathbb{H}^2)$ fixes a nonzero tangent vector then $\phi = \text{Id}_{\mathbb{H}^2}$. By the previous paragraph, we may assume the fixed vector is in $T_0 \mathbb{D}$. At $z = 0$ the hyperbolic norm is a scaling of the Euclidean norm, so the group preserving the norm and the orientation is $\text{SO}(2)$, which is rotations, so if some vector is fixed then the rotation is trivial. We know an isometry, $\text{Id}_{\mathbb{H}^2}$ that fixes 0 and whose derivative at 0 is the identity on $T_0 \mathbb{D}$. By Theorem 2.2.1, $\phi = \text{Id}_{\mathbb{H}^2}$. $\square$

COROLLARY 2.2.3 (Distance formula). For $z, w \in \mathbb{D}$:

$$d_{\mathbb{H}}(z, w) = 2 \operatorname{artanh} \left| \frac{z - w}{1 - \bar{w}z} \right| = 2 \operatorname{arsinh} \left(\frac{|z - w|}{\sqrt{(1 - |z|^2)(1 - |w|^2)}} \right)$$

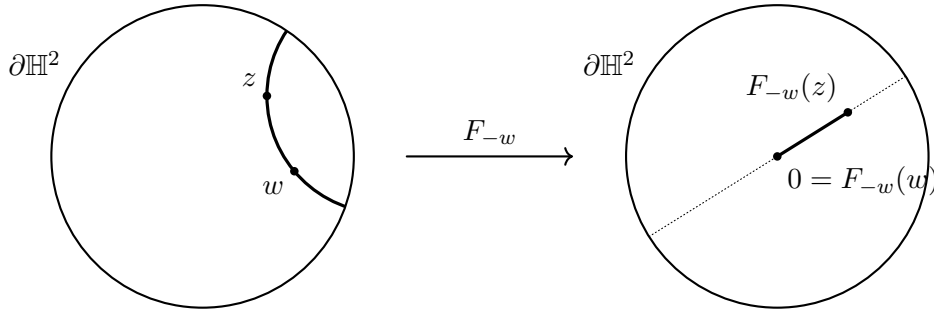
PROOF. The isometry F_{-w} takes w to 0. Apply Lemma 2.1.5 to $F_{-w}(z)$, using

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

2.3. Geodesics and the circle at infinity

PROPOSITION 2.3.1 (Geodesics in the disk). *Between any two points of $\mathbb{H}^2$ there is a unique geodesic, up to reparametrization. Every geodesic segment extends uniquely to a bi-infinite geodesic. Its image in the disk is either an arc of a Euclidean circle orthogonal to S^1 or a Euclidean diameter (which may be thought of as an arc of a “generalized circle”, Euclidean circle of infinite radius, orthogonal to S^1).*

PROOF. Given $z, w \in \mathbb{D}$, move w to 0 by an isometry F_{-w} , take the radial line from 0 to $F_{-w}(z)$, which is a geodesic segment, by Proposition 2.1.4, then move it back with F_w . This also gives uniqueness and extension to a complete line. Möbius maps send generalized circles to generalized circles, and are conformal: they preserve angles. Their images of diameters are therefore generalized circles that meet the unit circle orthogonally. $\square$



The circle S^1 is not part of the metric space: it is infinitely far away. But every geodesic ray converges to a point of S^1 . As in Lecture 1, we think of these points as *directions at infinity*.

We call two geodesic rays *asymptotic* if their images have finite Hausdorff distance. In a tree, this agrees with the definition from Lecture 1: two rays are asymptotic if they eventually coincide.

PROPOSITION 2.3.2 (Boundary of the hyperbolic plane). *Two geodesic rays in $\mathbb{H}^2$ are asymptotic if and only if their limiting endpoints on S^1 agree. Thus:*

$$\partial\mathbb{H}^2 \cong S^1$$

Every pair of distinct points of $\partial\mathbb{H}^2$ is joined by a unique bi-infinite geodesic.

PROOF. Suppose that geodesic rays α and β have the same endpoint in S^1 . Up to isometry, we may assume that $\alpha = [0, 1)$. If β lies on the real diameter, the conclusion is immediate. Otherwise, β is a subray of a Euclidean circle with center $(1, c)$ for some $c \neq 0$.

The circle containing β has equation:

$$(x - 1)^2 + y^2 = 2cy$$

For $z = x + iy \in \beta$ approaching 1, this gives $y = O((1 - x)^2)$. By Corollary 2.2.3:

$$\sinh\left(\frac{d_{\mathbb{H}}(z, x)}{2}\right) = \frac{|y|}{\sqrt{(1 - x^2 - y^2)(1 - x^2)}} = O(1 - x) \rightarrow 0$$

Thus the two rays approach one another, and in particular have finite Hausdorff distance.

Conversely, suppose α and β have finite Hausdorff distance, say at most R . Choose $z_n \in \alpha$ converging to its endpoint $\xi \in S^1$, and $w_n \in \beta$ with $d_{\mathbb{H}}(z_n, w_n) \leq R$. By Corollary 2.2.3:

$$|z_n - w_n| \leq 2 \sinh(R/2) \sqrt{(1 - |z_n|^2)(1 - |w_n|^2)} \rightarrow 0$$

Hence $w_n \rightarrow \xi$, so β has endpoint ξ as well.

Finally, two distinct points of S^1 determine a unique generalized Euclidean circle orthogonal to S^1 , which gives the unique bi-infinite geodesic joining them. $\square$

2.4. Classification of isometries

THEOREM 2.4.1 (Classification of orientation-preserving isometries). *Every nonidentity orientation-preserving isometry of $\mathbb{H}^2$ is exactly one of the following:*

Elliptic: *It fixes a unique point in $\mathbb{H}^2$. It is conjugate to a nontrivial rotation $z \mapsto e^{i\theta} z$ of the disk.*

Parabolic: *It has no fixed point in $\mathbb{H}^2$ and exactly one fixed point in $\partial\mathbb{H}^2$. It is conjugate to*

$$p_s(z) = \frac{(1 + is)z - is}{isz + 1 - is} \quad (s \in \mathbb{R} \setminus \{0\})$$

which fixes $1 \in S^1$.

Hyperbolic: *It has no fixed point in $\mathbb{H}^2$ and exactly two fixed points in $\partial\mathbb{H}^2$. It is conjugate to*

$$h_t(z) = F_{\tanh(t/2)}(z) = \frac{z + \tanh(t/2)}{1 + z \tanh(t/2)} \quad (t \in \mathbb{R} \setminus \{0\})$$

which fixes -1 and 1 and translates along the real diameter by signed distance t .

PROOF. By Proposition 2.2.2, every orientation-preserving isometry is represented by a matrix

$$A = \begin{pmatrix} a & b \\ \bar{b} & \bar{a} \end{pmatrix} \in SU(1, 1)$$

The fixed points of the corresponding Mobius transformation satisfy:

$$\bar{b}z^2 + (\bar{a} - a)z - b = 0$$

The discriminant is:

$$\Delta = (\bar{a} - a)^2 + 4|b|^2 = 4((\operatorname{Re} a)^2 - 1)$$

If $\Delta < 0$, there is exactly one fixed point in $\mathbb{D}$ (and one in $\mathbb{C} \cup \{\infty\} \setminus \bar{\mathbb{D}}$). If $\Delta = 0$, there is exactly one fixed point on S^1 . If $\Delta > 0$, there are exactly two fixed points on S^1 . (The case $b = 0$ gives a rotation directly, and the identity is excluded.)

In the $\Delta < 0$ case, conjugate the fixed point to 0. The resulting isometry belongs to $SO(2)$, so it is a rotation.

In the $\Delta > 0$ case, conjugate the geodesic joining the two fixed points to the real diameter. The resulting isometry fixes -1 and 1 , and therefore acts by translation on that diameter. It is consequently one of the transformations h_t , as follows from simple transitivity on unit tangent vectors.

In the remaining $\Delta = 0$ case, conjugate the unique boundary fixed point to 1 by a rotation. Since $\Delta = 0$, we may replace A by $-A$ if necessary and write $a = 1 + is$ for some real s . The conditions $|a|^2 - |b|^2 = 1$ and that 1 is fixed imply $b = -is$. Thus the transformation is p_s . $\square$

LEMMA 2.4.2 (Translation length and classification). *For an isometry ϕ of a metric space X , define its translation length by:*

$$\tau(\phi) := \inf_{x \in X} d(x, \phi(x))$$

Let ϕ be a nonidentity orientation-preserving isometry of $\mathbb{H}^2$. Then:

- (1) *ϕ is elliptic if and only if $\tau(\phi) = 0$ and the infimum is attained.*
- (2) *ϕ is parabolic if and only if $\tau(\phi) = 0$ and the infimum is not attained.*
- (3) *ϕ is loxodromic (hyperbolic) if and only if $\tau(\phi) > 0$. In this case the infimum is attained precisely on the axis of ϕ .*

PROOF. Translation length is invariant under conjugation, so by Theorem 2.4.1 it suffices to consider the standard representatives.

An elliptic isometry fixes a point, so its translation length is 0 and the infimum is attained.

For the parabolic isometry p_s , the distance formula gives:

$$\sinh\left(\frac{d_{\mathbb{H}}(z, p_s z)}{2}\right) = \frac{|s||1 - z|^2}{1 - |z|^2}$$

In particular, for $0 \leq r < 1$:

$$\sinh\left(\frac{d_{\mathbb{H}}(r, p_s(r))}{2}\right) = |s| \frac{1-r}{1+r} \rightarrow 0$$

as $r \rightarrow 1$. Thus $\tau(p_s) = 0$, but the infimum is not attained since p_s has no fixed point in $\mathbb{H}^2$.

Finally, consider the hyperbolic isometry $h_t = F_c$, where $c = \tanh(t/2) \in (-1, 1) \setminus \{0\}$. By the distance formula:

$$\sinh\left(\frac{d_{\mathbb{H}}(z, h_t z)}{2}\right) = \frac{|c||1-z^2|}{\sqrt{1-c^2}(1-|z|^2)} \geq \frac{|c|}{\sqrt{1-c^2}} = \sinh\left(\frac{|t|}{2}\right)$$

Equality holds precisely when z lies on the real diameter, which is the axis of h_t . Thus $\tau(h_t) = |t| > 0$. $\square$

LEMMA 2.4.3 (North–south dynamics). *Let ϕ be a loxodromic isometry of $\mathbb{H}^2$. Then ϕ has two fixed points $\xi^+, \xi^- \in \partial\mathbb{H}^2$ such that:*

$$\phi^n(\xi) \rightarrow \xi^+ \quad \text{for all } \xi \in \partial\mathbb{H}^2 \setminus \{\xi^-\}$$

and:

$$\phi^{-n}(\xi) \rightarrow \xi^- \quad \text{for all } \xi \in \partial\mathbb{H}^2 \setminus \{\xi^+\}$$

Both convergences are uniform on compact subsets of the indicated domains.

PROOF. By Theorem 2.4.1, up to conjugation we may assume $\phi = h_t$ for some $t > 0$. Its boundary fixed points are $\xi^+ = 1$ and $\xi^- = -1$.

Recall that:

$$h_t^n = F_{\tanh(nt/2)}$$

Writing $c_n = \tanh(nt/2)$, we have $c_n \rightarrow 1$ and:

$$h_t^n(z) - 1 = \frac{(1-c_n)(z-1)}{1+c_n z}$$

This tends to 0 uniformly on compact subsets of $S^1 \setminus \{-1\}$. Thus $h_t^n(\xi) \rightarrow 1$ uniformly on such subsets.

Similarly, $h_t^{-n} = F_{-c_n}$, and:

$$h_t^{-n}(z) + 1 = \frac{(1-c_n)(z+1)}{1-c_n z}$$

which tends to 0 uniformly on compact subsets of $S^1 \setminus \{1\}$. $\square$

Orientation-reversing isometries behave differently. For example, complex conjugation $z \mapsto \bar{z}$ is reflection in the real diameter: it fixes that geodesic pointwise and exchanges its two sides. Conjugating by isometries gives reflections in arbitrary geodesics.

Composing reflection in the real diameter with a nontrivial translation along that diameter gives a *glide reflection*:

$$g_t(z) = h_t(\bar{z}) \quad (t \neq 0)$$

It preserves the real diameter and translates along it, but has no fixed point in $\mathbb{H}^2$. Moreover, $g_t^2 = h_{2t}$. Every orientation-reversing isometry is either a reflection or a glide reflection, Exercise 2.2.

2.5. Hyperbolic surfaces and surface groups

PROPOSITION 2.5.1 (Regular hyperbolic polygons). *For every $n \geq 3$ and every angle*

$$0 < \alpha < \frac{n-2}{n}\pi$$

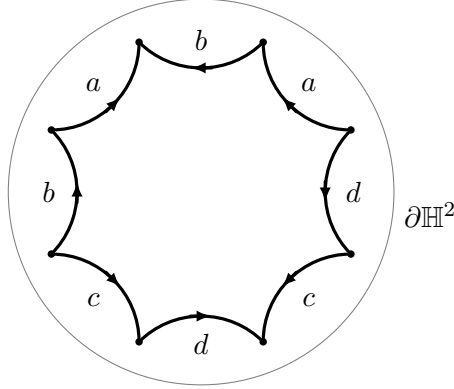
there exists a regular geodesic n -gon in $\mathbb{H}^2$ whose interior angles are all α .

IDEA OF THE PROOF. Place n points at equal angular intervals on a Euclidean circle of radius $r < 1$ centered at 0, and join consecutive points by hyperbolic geodesics. The resulting polygon is regular by rotational symmetry. Its interior angle depends continuously on r . As $r \rightarrow 0$ the angle tends to the Euclidean value $(n-2)\pi/n$; as $r \rightarrow 1$ it tends to 0. The intermediate value theorem completes the argument. $\square$

Choose a regular hyperbolic octagon P with interior angles $\pi/4$. Label its oriented sides, in cyclic order, by the word:

$$aba^{-1}b^{-1}cdc^{-1}d^{-1}$$

Join each side to the side bearing its inverse label, reversing boundary orientations. The label tells us which two sides are paired; the pairing is realized by an orientation-preserving hyperbolic isometry.



PROPOSITION 2.5.2 (The octagon surface). *The quotient obtained by identifying the sides of P is a closed, orientable surface S_2 of genus two. Moreover, the side-pairing isometries generate a group $\Gamma_2 \leq \text{Isom}^+(\mathbb{H}^2)$ with P as a fundamental polygon and:*

$$\mathbb{H}^2/\Gamma_2 \cong S_2$$

The action is free, properly discontinuous, and cocompact.

EXPLANATION. Under the side identifications all eight vertices become a single vertex. Their total angle is:

$$8 \cdot \frac{\pi}{4} = 2\pi$$

There is thus no cone singularity at the identified vertex. The quotient is orientable, with one vertex, four edges, and one face. Its Euler characteristic is:

$$\chi = 1 - 4 + 1 = -2 = 2 - 2g$$

so it is a closed orientable surface of genus $g = 2$.

The assertion that the side-pairings give a tiling of the entire hyperbolic plane, and that their group has P as fundamental polygon, is a special case of the *Poincaré polygon theorem*, which we use without proof. $\square$

The disk is simply connected, so its quotient map to S_2 is the universal covering. The group Γ_2 is therefore the fundamental group of S_2 . Reading the boundary word of the octagon gives the familiar presentation:

$$\Gamma_2 \cong \pi_1(S_2) \cong \langle a, b, c, d \mid [a, b][c, d] = 1 \rangle$$

where $[a, b] = aba^{-1}b^{-1}$.

More generally, for any $g \geq 2$ choose a regular hyperbolic $4g$ -gon with interior angles $\pi/(2g)$ and identify sides according to the cyclic boundary word $[a_1, b_1] \cdots [a_g, b_g]$. This yields a closed orientable surface S_g and an action of:

$$\pi_1(S_g) \cong \left\langle a_1, b_1, \dots, a_g, b_g \mid \prod_{i=1}^g [a_i, b_i] = 1 \right\rangle$$

freely, properly discontinuously, and cocompactly on $\mathbb{H}^2$.

2.6. Exercises

EXERCISE 2.1. Show that the Cayley transform

$$C(z) = i \frac{1+z}{1-z}$$

takes the Poincaré disk metric to $(dx^2 + dy^2)/y^2$ on the upper half-plane.

EXERCISE 2.2. Let p be a parabolic orientation-preserving isometry of $\mathbb{H}^2$, and let $R(z) = \bar{z}$. Show that $R \circ p$ is either a reflection in a geodesic or a glide reflection. Find examples where each possibility occurs.

EXERCISE 2.3. Check directly from the edge pairing of the octagon that all its vertices become a single vertex. Why does it matter that the total angle there is 2π ?

EXERCISE 2.4. For the $4g$ -gon construction, compute the Euler characteristic of the quotient from its numbers of vertices, edges, and faces.

EXERCISE 2.5. Why can no nonidentity element of the surface group Γ_2 act as an elliptic isometry?

EXERCISE 2.6 (Geodesics and tangent vectors). Let $p \in \mathbb{H}^2$ and let $0 \neq v \in T_p \mathbb{H}^2$. Show that, up to orientation-preserving reparametrization, there is a unique complete geodesic through p whose tangent vector at p is a positive multiple of v .

EXERCISE 2.7 (Angles of hyperbolic triangles). **Gauss–Bonnet theorem:** If D is a compact region in an oriented Riemannian surface with piecewise smooth boundary, then:

$$\int_D K \, dA + \int_{\partial D} k_g \, ds + \sum_{j=1}^n (\pi - \theta_j) = 2\pi \chi(D)$$

where K is the Gaussian curvature, k_g is the geodesic curvature of the smooth parts of the boundary, θ_j are the interior angles at the corners, and $\chi(D)$ is the Euler characteristic.

The hyperbolic plane $\mathbb{H}^2$ has constant curvature $K = -1$. The geodesic curvature of a geodesic is zero. Let T be a nondegenerate hyperbolic geodesic triangle with interior angles α , β , and γ . Its Euler characteristic is 1.

(1) Use Gauss–Bonnet to prove that:

$$\text{Area}(T) = \pi - \alpha - \beta - \gamma$$

Deduce that:

$$\alpha + \beta + \gamma < \pi$$

(2) Conversely, suppose $\alpha, \beta, \gamma \in (0, \pi)$ satisfy $\alpha + \beta + \gamma < \pi$. Prove that there exists a hyperbolic geodesic triangle with these angles.

(3) Prove that this triangle is unique up to hyperbolic isometry.

EXERCISE 2.8 (Area of closed hyperbolic surfaces). Use the Gauss–Bonnet formula for a geodesic n -gon with interior angles $\alpha_1, \dots, \alpha_n$:

$$\text{Area}(P) = (n-2)\pi - \sum_{j=1}^n \alpha_j$$

to compute the area of the regular hyperbolic $4g$ -gon used in the construction of S_g . Deduce that:

$$\text{Area}(S_g) = 4\pi(g-1)$$

EXERCISE 2.9 (Ideal triangles and thin triangles). An *ideal triangle* in $\mathbb{H}^2$ is the closed convex region bounded by three bi-infinite hyperbolic geodesics joining three distinct points of $\partial \mathbb{H}^2$.

A geodesic triangle (finite or ideal) is called δ -thin if every point on each side lies at distance at most δ from the union of the other two sides.

- (1) Show that any two ideal triangles are related by an isometry of $\mathbb{H}^2$.
- (2) Let T be a finite geodesic triangle in the Poincaré disk. Show that some hyperbolic isometry takes one side of T into the real diameter and its opposite vertex onto the positive imaginary axis. Deduce that the image of T is contained in the ideal triangle Δ with vertices -1 , 1 , and i .
- (3) Show that Δ is δ -thin for:

$$\delta = \operatorname{arsinh}(1) = \log(1 + \sqrt{2})$$

Hint: Let g_+ and g_- be the sides of Δ joining i to 1 and -1 , respectively. Show that for $x \in (-1, 1)$:

$$d_{\mathbb{H}}(x, g_+) = \operatorname{arsinh}\left(\frac{1-x}{1+x}\right)$$

$$d_{\mathbb{H}}(x, g_-) = \operatorname{arsinh}\left(\frac{1+x}{1-x}\right)$$

To compute these distances, use the disk isometry:

$$z \mapsto \frac{z-x}{1-xz}$$

which sends x to 0 . Use the symmetries of Δ to deal with the other two sides.

- (4) Deduce that every finite geodesic triangle in $\mathbb{H}^2$ is δ -thin, with the same constant δ .

Hint: Given a point p on one side of a finite triangle, normalize the triangle as in part (b). Choose a point q on one of the other two sides of the enclosing ideal triangle with $d_{\mathbb{H}}(p, q) \leq \delta$. Explain why the geodesic segment $[p, q]$ must meet one of the other two sides of the finite triangle.

EXERCISE 2.10 (Area, circumference, and isoperimetry). Let B_R be the hyperbolic ball of radius R centered at 0 in the Poincaré disk. Recall that B_R is a Euclidean disk of radius $\tanh(R/2)$.

- (1) Using the hyperbolic metric

$$ds = \frac{2|dz|}{1-|z|^2}$$

compute the hyperbolic circumference $L(R)$ of ∂B_R . Show that:

$$L(R) = 2\pi \sinh R$$

- (2) Using the hyperbolic area element

$$dA = \frac{4 \, dx \, dy}{(1-|z|^2)^2}$$

compute the hyperbolic area $A(R)$ of B_R . Show that:

$$A(R) = 2\pi(\cosh R - 1)$$

- (3) Show that:

$$L(R)^2 = A(R)^2 + 4\pi A(R)$$

Compare this with the corresponding formula for Euclidean disks.

- (4) Show that $A(R) \leq L(R)$ for all $R > 0$, and that:

$$\lim_{R \rightarrow \infty} \frac{L(R)}{A(R)} = 1$$

Explain how this differs from the behavior of Euclidean disks.