Geometric group theory

Lawk Mineh

Contents

1. Preface

These notes follow a course in geometric group theory I taught in the winter semester of 2025-26 at the University of Bonn. The course was intended as a introduction to the field, and is aimed primarily at students who have some background in algebra, topology, and geometry. There are relatively few prerequisites for this course: a thorough understanding of the fundamentals of group theory (groups, subgroups, normality, group actions) and basic topology (compactness, separation axioms, metric spaces) and algebraic topology (fundamental groups, simplicial complexes) are required, but not much more is necessary. Some familiarity with covering spaces is also recommended, and ideas from the theory of manifolds may prove helpful.

We will begin by introducing the reader to the key concepts underpinning the coarse geometry of groups, before moving to hyperbolic metric spaces and the groups that act nicely on them. The topic of *hyperbolic groups*, now becoming classical, is at once rich and detailed enough that it allows one to touch on many of the salient areas of geometric group theory at once, while also basic enough that it may be presented as an accessible primer on the subject as a whole to an advanced undergraduate or graduate audience. We will develop enough theory to be able to prove many of the most important fundamental theorems about hyperbolic groups. At the end of the course, we will also see some groups acting on trees — Bass–Serre theory — without which I felt an introduction to geometric group theory could not be entirely complete.

I have chosen the route of understanding hyperbolic groups mostly using dynamics, through the framework of *convergence groups*. While this does take some time to develop, it is both a very aesthetically appealing and powerful approach, and I feel as though it is somewhat under-represented in the current literature on the topic. The material on this is somewhat spread out across many sources (some of which are quite difficult to find!), and to my knowledge there is not as of yet a unified account of the theory. I attempt to rectify this somewhat here, though the constraints on the nature of the course obviously yield some restrictions. As such, I hope that even some graduate students and some researchers in the field who are less familiar with this area may find this helpful.

CHAPTER 1

Fundamentals

1. Group presentations

We will begin with the notion of a group presentation, which is a fundamental way to express an abstract (discrete) group. A group presentation is a description of a group in terms of 'generators' and 'relations'. That is, some free variables and the equations bounding them. To make this notion precise, we recall the free group.

DEFINITION 1.1 (Free group). Let S be a set. The *free group generated by* S is the group $F(S)$ such that for any group G and function $f\colon S \to G$, there is a unique homomorphism $\widehat{f}\colon F(S) \to G$ making the following diagram commute

$$\begin{tikzcd} F(S) \arrow[dr, dashed, "\widehat{f}"] & \\ S \arrow[u, hook] \arrow[r, "f"'] & G \end{tikzcd}$$

where $S \to F(S)$ is a natural inclusion.

In other words, a free group on a set S is the image of S in **Grp** under the free functor **Set** $\to$ **Grp**.

EXERCISE 1.2. Show that $F(S)$ and $F(T)$ are isomorphic if and only if S and T are in bijection. It follows that a free group is uniquely determined up to isomorphism by the cardinality of its generating set.

DEFINITION 1.3 (Rank of a free group). Let S be a set. The *rank* of $F(S)$ is the cardinality $|S|$ of S.

It is often useful to have a practical model of the free group that one can refer to, when it is unwieldy or otherwise not possible to use the definition above in terms of a universal property.

DEFINITION 1.4. Let S be a set. Denote by S^{-1} the set in bijection with S, whose elements are the symbols s^{-1} for each $s \in S$: these are the formal inverses of elements of S. We identify this bijection $\cdot^{-1}\colon S \to S^{-1}$ with its inverse, so that we may write $(s^{-1})^{-1} = s$. A *word in* S is an ordered finite sequence of elements in $S \cup S^{-1}$; the *empty word* is the empty sequence. The *length* $\ell(w)$ of a word w is the number of terms in the sequence. A word is called *reduced* if it contains no consecutive terms of the form ss^{-1} for $s \in S \cup S^{-1}$. Define the equivalence relation $=_{F(S)}$ on words as the symmetric and transitive closure of deleting such an element-inverse pair. Note that any word is equivalent to a reduced word.

The free group $F(S)$ on S is the set of words in S up to the above equivalence relation, with the operation of concatenation of (class representatives) of words. It is straightforward to check that this is a well-defined operation. The identity of this group is the empty word.

It will usually not cause confusion for us to identify words in S with their equivalence classes up to reduction, so we will interchangeable refer to words as 'being' elements of $F(S)$ as well as 'representing' elements of $F(S)$. Note that every group is the quotient of a free group: indeed, if G is a group, then applying the forgetful functor it can be viewed as a set, and the universal property implies there is a unique homomorphism $F(G) \to G$ acting as 'evaluation' of words in G.

Definition 1.5 (Group presentation). Let S be a set, and R a set of words in S. Let G be the quotient group $F(S)/\langle\!\langle R \rangle\!\rangle$ and write $\langle S \,|\, R\rangle$ for the *presentation of* G with *generators* S and *relators* R. If both S and R are finite sets, then the presentation is called *finite*. If a word w in S represents an element $g \in G$, we may write $w =_G g$.

We say that G is *finitely generated* if it admits a presentation $\langle S \,|\, R\rangle$ with S finite. Equivalently, it is the quotient of a finite rank free group. Further, G is *finitely presented* if it admits a finite presentation.

Exercise 1.6. Show that the group of invertible $n \times n$ integer matrices $\mathrm{GL}_n(\mathbb{Z})$ is finitely generated. Show that the rational numbers $\mathbb{Q}$ do not form a finitely generated group under addition.

Exercise 1.7. Show that the properties of being finitely generated and being finitely presented are stable under extensions.

When we have explicit sets to work with, we often write a group presentation with the elements of S and R, omitting the set brackets.

Example 1.8.

(1) The free group on S has a presentation with generating set S and no relators;
(2) the free abelian group on S has a presentation with generators S and commutators $[s, t]$ as relators for each $s, t \in S$;
(3) the fundamental group of the genus g surface Σ has presentation
$$\langle a_1, \dots, a_g, b_1, \dots, b_g \,|\, [a_1, b_1] \dots [a_g, b_g]\rangle.$$
To see this, observe that Σ can be obtained as the quotient space of a $4g$-gon. One should convince oneself first about the case that Σ the torus $g = 1$, then observe that higher genus surfaces are obtained by taking connected sums of tori and lower genus surfaces. The $4g$-gon can be cut into g hexagons, each of which are tori with a single boundary component under the edge identifications;
(4) a cyclic group of order n has presentation $\langle a \,|\, a^n\rangle$.

Of course, a presentation does not determine a group uniquely. Indeed, though some information can be gleaned from a presentation in specific circumstances, group presentations in general do not encode readily accessible information about a group. However, if we know which group we are working with to start off with, two of its presentations are not *too* unrelated (at least, when it comes to finite presentations).

DEFINITION 1.9 (Tietze transformations). Let $\mathcal{P} = \langle S \mid R \rangle$ be a presentation for a group G. The following four operations are *Tietze transformations*, taking the presentation $\mathcal{P}$ to a presentation $\mathcal{P}'$:

(i) Let $r \in \langle\!\langle R \rangle\!\rangle$ be a word in $F(S)$. Define $\mathcal{P}' = \langle S \mid R \cup \{r\} \rangle$.
(ii) Suppose $r \in R$ is such that $r \in \langle\!\langle R - \{r\} \rangle\!\rangle$. Define $\mathcal{P}' = \langle S \mid R - \{r\} \rangle$.
(iii) Let t be an element in $F(S)$ and w a word in S representing t. Define $\mathcal{P}' = \langle S \cup \{t\} \mid R \cup \{t^{-1}w\} \rangle$.
(iv) Suppose $s \in S$ is such that s can be written as a word w in $S' \subset S$, and $s^{-1}w \in R$. Define $\mathcal{P}' = \langle S - \{s\} \mid R - \{s^{-1}w\} \rangle$.

These operations correspond to adding a superfluous relator, deleting a superfluous relator, adding a superfluous generator, and deleting a superfluous generator respectively.

It is a tedious, though possibly instructive, exercise to verify that each of the Tietze transformations preserve the isomorphism type of the presented group G. We now have the important observation about finite presentations.

LEMMA 1.10 (Tietze's theorem). *Let $\mathcal{P}$ and $\mathcal{P}'$ be two finite presentations of a given group G. Then there exist a finite sequence of Tietze transformations that transform $\mathcal{P}$ into $\mathcal{P}'$.*

PROOF. The idea is that one can arrive at a common presentation for both $\mathcal{P}$ and $\mathcal{P}'$ by adding in all of the generators and then relators from both presentations in one at a time. We leave the details to the reader. $\square$

2. Groups and their actions

Recall that an *action* of a group G on a set X is a function $\cdot\colon G \times X \to X$ satisfying

- $1 \cdot x = x$ for all $x \in X$; and
- $g \cdot (h \cdot x) = (gh) \cdot x$ for all $g, h \in G$ and $x \in X$.

Let us introduce some terminology for group actions.

DEFINITION 2.1 (Group actions). Let X be a topological space equipped with a G-action $\cdot\colon G \times X \to X$. We say the action is:

- *cocompact* if the quotient space X/G is compact, with the quotient topology;
- *properly discontinuous* if for any compact $K \subseteq X$, the set
$$\{g \in G \mid gK \cap K \neq \emptyset\}$$
is finite;

and when X is a metric space, with metric d:

- *by isometries* if for any $g \in G$ and $x, y \in X$, we have
$$\mathrm{d}(g \cdot x, g \cdot y) = \mathrm{d}(x, y)$$
- *geometric* if it is by isometries, cocompact, and properly discontinuous.

It is often useful to think about an action of a group G on a set X as a homomorphism $G \to \mathrm{Aut}(X)$ from the group to the automorphisms of the set. We say automorphisms here, rather than permutations, for X is often endowed with some structure, and the action is required to preserve that structure. For instance, if X is a topological space, we want the group elements to act by homeomorphisms, when X is a metric space, usually by isometries, and so on.

We now introduce the most basic geometric-combinatorial object for our study: *Cayley graphs* of groups. For us, a *graph* Γ will be a set of *vertices* $V\Gamma$ and a set of *edges* $E\Gamma$, which comes equipped with a pair of functions $\iota, \tau\colon E\Gamma \to V\Gamma$ (denoting the *initial* and *terminal* endpoints of an edge). We will say that two vertices $v, w \in V\Gamma$ are *connected* by an edge $e \in E\Gamma$ with $\iota(e) = v$ and $\tau(e) = w$, and we write $v \sim w$ in this case.

The *geometric realisation of a graph* Γ is a simplicial complex whose 0-skeleton is $V\Gamma$, and whose 1-simplices are the edges $E\Gamma$, with attaching maps determined by the incidence functions ι and τ. We equip this complex with the metric induced by giving each edge unit length. Throughout, we will identify a graph with its geometric realisation.

Definition 2.2 (Cayley graph). Let G be a group with generating set S. The *Cayley graph* of G with respect to S is the graph $\Gamma(G, S)$ whose vertex set is G, and with an edge $g \sim h$ if there is $s \in S$ with $gs = h$.

Note that G acts (by left multiplication) transitively on the vertex set of $\Gamma(G, S)$, and with $|S|$-many orbits of edges. It is straightforward to see that this is an isometric action. When the group G is finitely generated and S is a finite set, there are finitely many edge orbits, and so the action is also properly discontinuous and cocompact. This gives us our archetypal model for a geometric action. We will see later that essentially every geometric action of a finitely generated group is like one on a Cayley graph.

Definition 2.3 (Word metric). Let G be a group and S be a generating set. The *word metric* on G with respect to S is the metric d_S defined as

$$\mathrm{d}_S(g, h) = \min\{\ell(w) \mid w =_G g^{-1}h\}.$$

We will write $|g|_S = \mathrm{d}_S(1, g)$ for the *length* of g with respect to S.

The word metric on a group coincides with the restriction of the edge-path metric on the associated Cayley graph to its vertex set, and the length of an element with respect to a generating set is exactly the length of the shortest word representing that element in that generating set.

Example 2.4. The free abelian group $\mathbb{Z}^n = \langle a_1, \ldots, a_n \mid [a_i, a_j] = 1\rangle$ acts geometrically on the Euclidean space $\mathbb{R}^n$ of dimension n, by the translations

$$a_i \cdot (x_1, \ldots, x_n) = (x_1, \ldots, x_i + 1, \ldots, x_n).$$

This is, by construction, an action by isometries, and one should check that the action is properly discontinuous. The quotient space $\mathbb{R}^n/\mathbb{Z}^n = \mathbb{T}^n$ is the Euclidean n-torus, the product of n copies of the circle S^1. One can view this as the covering space action of $\pi_1(\mathbb{T}^n) \cong \mathbb{Z}^n$ on its universal cover $\mathbb{R}^n$.

Taking $S = \{a_1, \ldots, a_n\}$ as the standard generating set as in the presentation above. The Cayley graph $\Gamma = \Gamma(\mathbb{Z}^n, S)$ embeds as the integer grid in $\mathbb{R}^n$, and the action of $\mathbb{Z}^n$ above restricts to the standard action on Γ by left multiplication. Note that this embedding of Γ into $\mathbb{R}^n$ is not quite isometric: one can show it is distance non-increasing, and decreases distances by at most a factor of $\sqrt{n}$.

Example 2.5. Let S be a set, and $F = F(S)$ the free group on S. The Cayley graph $\Gamma(F, S)$ is the regular tree with valence $2|S|$: each vertex has an outgoing edge labelled s and an incoming edge with labelled s^{-1}. Note that F is the fundamental group of the wedge of $|S|$ circles, whose universal cover is $\Gamma(F, S)$.

Example 2.6 (Cyclic group). An infinite cyclic group $\langle a \rangle \cong \mathbb{Z}$ is a free group, and has an obvious one-generator presentation with no relators. As a metric space, the Cayley graph with respect to this generating set is, of course, just the line $\mathbb{R}$. Consider, however, the generating set $S = \{a^2, a^3\}$ for $\langle a \rangle$. The Cayley graph $\Gamma = \Gamma(\mathbb{Z}, S)$ is definitely not a line: it has many loops for instance.

However, the map $\Gamma \to \mathbb{R}$ defined by taking the identity on $V\Gamma = \mathbb{Z}$ and mapping each edge to its numerically lesser endpoint distorts distances additively by at most 3. In this way, Γ is 'coarsely' isometric to the real line.

Exercise 2.7. Draw a portion of the Cayley graph for the group with presentation

$$\langle a, t \,|\, tat^{-1} = a^2 \rangle.$$

This is an example of a *Baumslag-Solitar group*: these form a two-parameter family of groups, indexed by integers $m, n \in \mathbb{Z}$, with a relator $ta^m t^{-1} = a^n$.

3. Quasi-isometries

The examples of the previous section serve to illustrate a key point: though the exact metric on two spaces with a geometric G-action may differ on a local level, the large-scale 'rough' geometry of the spaces remains the same. The takeaway is that, for our purposes, isometry is not the correct notion of morphism, motivating the following.

Definition 3.1 (Quasi-isometry). Let X and Y be metric spaces, $\lambda \geq 1$, and $c \geq 0$. A map $f\colon X \to Y$ is a (λ, c)-*quasi-isometric embedding*

$$\frac{1}{\lambda}\,\mathrm{d}_X(x, x') - c \leq \mathrm{d}_Y(f(x), f(x')) \leq \lambda\,\mathrm{d}_X(x, x') + c$$

for all $x, x' \in X$.

Moreover, f is K-*coarsely surjective* if for every $y \in Y$, there is $x \in X$ with $\mathrm{d}_Y(f(x), y)$. A (λ, c)-*quasi-isometry* is a (λ, c)-quasi-isometric embedding that is K-coarsely surjective for some $K \geq 0$.

As with many definitions we will give, we will omit the constants when they are not important to the discussion. That is, for example, we say a map f is simply a *quasi-isometry* if there are some $\lambda \geq 1$ and $c \geq 0$ such that it is a (λ, c)-quasi-isometry. It is important to note that a quasi-isometry need not be continuous!

Exercise 3.2. Show that if $f\colon X \to Y$ is a quasi-isometry, it has a *quasi-inverse*: a quasi-isometry $g\colon Y \to X$ such that $g \circ f$ and $f \circ g$ are a finite distance from the identity on X and Y respectively (with respect to the supremal metric on functions), in such a way that all the the constants involved depend only on those of f (i.e. they are independent of the particular function).

Show that quasi-isometry is an equivalence relation on metric spaces.

Example 3.3. Given $n \geq 2$, write T_n for the n-regular tree. For any $m, n \geq 3$, the trees T_m and T_n are quasi-isometric. Transitivity of quasi-isometry means that it suffices to show that any n-regular tree is quasi-isometric to T_3. The key fact is that if one collapses an edge of T_3, then one combines two vertices and increases the valence by one.

Thus, let us take a spanning forest $\mathfrak{T}$ of T_3 by disjoint paths of length $n-3$ (note: one needs the axiom of countable choice for this), and consider the map $f\colon T_3 \to T_3/\mathfrak{T} \cong T_n$ obtained by collapsing each connected component of $\mathfrak{T}$ to a single point. The map f is of course distance non-increasing. Moreover, at least every $(n-2)^{\text{th}}$ edge of an arc in T_3 must lie outside of $\mathfrak{T}$, so that

$$\mathrm{d}_{T_3}(x,y) \leq (n-2)\,\mathrm{d}_{T_n}(f(x),f(y)) + (n-3)$$

for any $x, y \in T_3$. Finally, our map is surjective, so it is a quasi-isometry.

Example 3.4. We will give a sketch that the real line $\mathbb{R}$ and the ray $[0,\infty)$ are not quasi-isometric, and leave it to the reader to assemble the details. Indeed, suppose that $\varphi\colon \mathbb{R} \to [0,\infty)$ is a quasi-isometric embedding, and suppose for simplicity that $\varphi(0) = 0$. As $t \to \infty$, we must have $\varphi(t) \to \infty$ and $\varphi(-t) \to \infty$, since φ coarsely preserves distances. Pick some very large $x \in [0,\infty)$. There must then be some correspondingly very large $s, t > 0$ so that $\varphi(-s)$ and $\varphi(t)$ lie a uniformly bounded distance from x. Using the fact that φ is a quasi-isometric embedding, $t - (-s) = t + s$ is uniformly bounded. Provided one picks x large enough, then $t + s$ may take arbitrarily large values: a contradiction.

Exercise 3.5.

(1) Show that the half-plane $\mathbb{R}^2_+ = \{(a,b) \in \mathbb{R}^2 \mid b > 0\}$ is not quasi-isometric to the whole plane $\mathbb{R}^2$.
(2) Show that $\mathbb{R}^n$ is not quasi-isometric to T_m for any $m, n \geq 2$.
(3) Show that if $\mathbb{R}^m$ is quasi-isometric to $\mathbb{R}^n$, then $m = n$. (Hint: Consider the volume growth rate of balls in $\mathbb{R}^n$. That is, the growth rate of the function $\mathrm{vol}_n\colon r \mapsto \mathrm{vol}(B_{\mathbb{R}^n}(0,r))$, which is polynomial in degree n. Show that this rate is — up to a suitable equivalence relation — preserved under quasi-isometries.)

One of the immediate upshots of this notion is that the quasi-isometry type of the Cayley graph of a finitely generated group is an invariant of the group.

Lemma 3.6. *Let S and T be finite generating sets for group G. Then (G, d_S) and (G, d_T) are quasi-isometric.*

Proof. It is enough to show that (G, d_S) is quasi-isometric to $(G, \mathrm{d}_{S\cup T})$. The result then follows by transitivity of quasi-isometry. We may thus suppose without loss of generality that $S \subset T$, so $S = \{s_1, \dots, s_n\}$ and $T = S \sqcup \{t_1, \dots, t_m\}$, with $m \geq 1$.

The identity map is our candidate for a quasi-isometry. Of course, this map is surjective. It is immediate that the identity is distance non-increasing: a word in S representing an element of G must be at least as long as one in T. As S is a generating set, each of the elements t_i can be expressed as a word w_i in S. Let $\lambda = \max\{\ell(w_i) \mid i = 1, \dots, m\} < \infty$ be the maximum over the lengths of these words.

Take any $g, h \in G$, and let w be a word of minimal length in T representing $g^{-1}h$. We may replace any instance of t_i in w with the word w_i to obtain a word w' in S representing $g^{-1}h$. Since we are, at worst, replacing each letter with λ letters, we have

$$\mathrm{d}_S(g,h) \leq \ell(w') \leq \lambda\ell(w) = \lambda\,\mathrm{d}_T(g,h).$$

It follows that the identity map is a $(\lambda, 0)$-quasi-isometry. $\square$

In light of the above, we will say from now on that a finitely generated group G is *quasi-isometric* to a space X if, when equipped with a word metric with respect to a finite generating set, it is quasi-isometric to X. The above lemma shows that this notion is well-defined up to change of finite generating set.

EXERCISE 3.7. Think about what the Cayley graph of a group looks like with respect to a generating set that gives a finite presentation. Come up with a graph-theoretic characterisation of a group being finitely presentable, and use this to show that finite presentability is a quasi-isometry invariant.

DEFINITION 3.8 (Quasi-geodesics). Let $\lambda \geq 1$ and $c \geq 0$. A (λ, c)-*quasi-geodesic* in X is a (λ, c)-quasi-isometric embedding of a closed interval $I \subseteq \mathbb{R}$ into X. The *endpoints* of a quasi-geodesic are the images of the endpoints of the interval, if the interval is bounded.

We call a $(1, 0)$-quasi-geodesic a *geodesic*. A space X is called a *geodesic space* if every pair of points can be joined by a geodesic. Given two points $x, y \in X$ we will often denote by $[x, y]$ a choice of geodesic whose endpoints are x and y.

REMARK 3.9. A $(\lambda, 0)$-quasi-geodesic is necessarily continuous. In particular, geodesics are continuous.

EXAMPLE 3.10. The plane $\mathbb{R}^2$ is a geodesic space, but $\mathbb{R}^2 - \{0\}$ is not.

Quasi-geodesics need not be continuous in general. The following lemma allows us to restrict our attention to continuous quasi-geodesics in many scenarios, without losing very much. We only sketch a very rough idea here and leave the details to the reader, though those less interested in this rather dull technical exercise may also find a proof in Bridson–Haefliger, Lemma III.H.1.11.

LEMMA 3.11. *Let X be a geodesic space, $\lambda \geq 1$, and $c \geq 0$. There is a constant $c' = c'(\lambda, c) \geq 0$ such that the following is true.*

Let $\gamma\colon I \to X$ be a (λ, c)-quasi-geodesic in X. Then there is a continuous (λ, c')-quasi-geodesic $\gamma'\colon J \to X$ with the same endpoints as γ, such that the images of γ and γ' are a Hausdorff distance of at most c' from one another.

PROOF. Partition I along its integer points, and construct γ' by concatenating geodesics joining the images of this partition in X. As γ is (λ, c)-quasi-geodesic, the length of each of these segments is at most $\lambda + c$, so the second claim holds as long as we choose c' greater than this. For the first and last claims, we can use the quasi-geodesicity of γ and compare the paths by their images of the integer partition. $\square$

Recall that the Heine–Borel theorem tells us that closed and bounded subsets of finite dimensional Euclidean spaces are compact. Many other spaces we will consider have this important property, such as locally finite graphs. We view this as a sort of finiteness property; non-proper spaces include things such as Banach spaces of infinite dimension and graphs with infinite valence at a vertex. We will give a name to metric spaces satisfying this property more generally.

DEFINITION 3.12 (Proper space). A metric space X is called *proper* if each of its closed and bounded sets are compact.

EXERCISE 3.13. Let X be a proper metric space, and suppose $G \leqslant \mathrm{Isom}(X)$ is a subgroup of isometries. Show that the action of G on X is properly discontinuous if and only if G is a discrete subgroup of $\mathrm{Isom}(X)$, equipped with the compact-open topology.

EXERCISE 3.14. A *length space* is a metric space where any two points can be joined by a rectifiable path, and the distance between two points coincides with the infimum of the lengths of all such paths. Prove the *Hopf–Rinow theorem*: every complete and locally compact length space is proper and geodesic.

Perhaps the most important observation we will make is the following, which is sometimes referred to as the *fundamental lemma of geometric group theory.*

PROPOSITION 3.15 (Milnor–Schwarz Lemma). *Let X be a proper geodesic metric space, and suppose that G acts on X cocompactly by isometries. Then there is a generating set S of G such that the orbit map*

$$G \to X, g \to g \cdot x$$

is a quasi-isometry for any $x \in X$. Moreover, if the action is properly discontinuous, then S is finite.

PROOF. Let $x \in X$ be an arbitrary point. As X is proper, it is locally compact. Therefore cocompactness of the action is equivalent to the existence of compact $B \subseteq X$ with $x \in B$ such that $G \cdot B = X$. As B is compact, it is a bounded set: let R be the diameter of B and define

$$S = \{s \in G \mid \mathrm{d}_X(x, s \cdot x) \leq 3R\} - \{1\},$$

We first show that S is a generating set for G. Let $g \in G$ be an element and write $d = \mathrm{d}_X(x, g \cdot x)$. As X is a geodesic space, there is a geodesic $\gamma\colon [0, d] \to X$ between x and $g \cdot x$. We may choose a partition $0 = t_0 < \cdots < t_n = d$ of $[0, d]$ such that $t_i - t_{i-1} = R$ for each $i = 1, \ldots, n-1$ and $t_n - t_{n-1} \leq R$. It follows that

$$n \leq \frac{1}{R}\mathrm{d}_X(x, g \cdot x) + 1. \tag{3.1}$$

As $G \cdot B = X$, there is $g_i \in G$ such that $\gamma(t_i) \in g_i B$ for each $i = 0, \ldots, n$. We may take $g_0 = 1$ and $g_n = g$.

For each $i = 1, \ldots, n,$, we have

$$\begin{aligned}\mathrm{d}_X(g_{i-1} \cdot x, g_i \cdot x) &\leq \mathrm{d}_X(g_{i-1} \cdot x, \gamma(t_{i-1})) + \mathrm{d}_X(\gamma(t_{i-1}), \gamma(t_i)) + \mathrm{d}_X(\gamma(t_i), g_i \cdot x)\\ &\leq R + R + R = 3R\end{aligned}$$

so that $\mathrm{d}_X(x, g_{i-1}^{-1} g_i \cdot x) \leq 3R$. It follows by definition that $g_{i-1}^{-1} g_i \in S$ for each $i = 1, \ldots, n$. By a finite induction it follows that $g = g_n \in \langle S \rangle$. As g was arbitrary, S generates G.

At this point we remark that since X is proper, then the ball of radius $3R$ about x is compact. Hence, if the action of G is properly discontinuous, then the set S is finite.

We now show that the map in the statement is a quasi-isometry. Since the map is G-equivariant, we need only bound $\mathrm{d}_X(x, g \cdot x)$ from above and below by linear functions of $\mathrm{d}_S(1, g)$. From the above, g can be written as a word of length n in S, namely

$$g = g_1(g_1^{-1} g_2)(g_2^{-1} g_3) \ldots (g_{n-2}^{-1} g_{n-1})(g_{n-1}^{-1} g_n).$$

Hence $\mathrm{d}_S(1,g) \leq n$. By (3.1), this implies $\mathrm{d}_S(1,g) \leq \frac{1}{R}\mathrm{d}_X(x, g\cdot x) + 1$. Moreover, if $w = s_1 \dots s_n$ is a word of minimal length representing g. Then

$$\begin{aligned}\mathrm{d}_X(x, g\cdot x) &\leq \mathrm{d}_X(x, s_1 x) + \cdots + \mathrm{d}_X(s_1 \dots s_{n-1}\cdot x, g\cdot x)\\ &\leq \sum_{i=1}^{n} \mathrm{d}(x, s_i \cdot x)\\ &\leq 3Rn = 3R\,\mathrm{d}_S(1,g)\end{aligned}$$

where the first inequality holds by the triangle inequality, the second by the fact G acts by isometries, and the third by the definition of S. □

Recall that two groups are *commensurate* if they contain isomorphic subgroups of finite index. We have a basic consequence of the Milnor–Schwarz lemma

LEMMA 3.16. *Finitely generated commensurate groups are quasi-isometric.*

PROOF. It is enough to show that a group is quasi-isometric to any of its finite index subgroups. Let $H \leqslant_f G$ and let S be a finite generating set for G. As G acts properly discontinuously by isometries on $\Gamma(G,S)$, so does H. Moreover, every point of $\Gamma(G,S)$ is at most $[G : H]$ from H, so the action is cocompact. Hence by the Milnor–Schwarz lemma, H has a finite generating set T for which $\Gamma(H,T)$ is quasi-isometric to $\Gamma(G,S)$. □

REMARK 3.17. Albert Schwarz, whose name appears as the second component of above named result, is a Russian-born mathematician who, after beginning in topology, spent a majority of his career working on mathematical physics. The name Schwarz is a German-Jewish name, and was transliterated to Russian as Шварц. Many sources still cite this result as the 'Švarc–Milnor' or 'Milnor–Švarc' lemma, owing to a curious decision by the AMS in the 1950s to re-transliterate Шварц as Švarc. Amusingly, Schwarz later moved to the United States, where he goes by the original spelling of his family name.

REMARK 3.18. The main initial motivation for considering quasi-isometries comes from differential geometry; they fundamentally clarify the relationship between continuous structures and certain discrete objects approximating them. In particular, Schwarz and Milnor were interested in relating volume growth in universal covers of Riemannian manifolds to some notion of growth in their fundamental groups. That these rates are the same for compact manifolds, up to a suitable equivalence relation, is a straightforward consequence of the Milnor–Schwarz lemma.

We conclude this section with the statement of a major theorem of Gromov. The proof is well beyond the scope of this course, but it is a strong indicator that one can recover a remarkable amount of algebraic information from asymptotic geometric data. We will not give a precise definition here, but the *growth rate* of a finitely group is the rate of growth of the function $r \mapsto |B(1,r)|$, where the ball is taken the group with respect to some word metric for a finite generating set. It is not difficult to see that this is a quasi-isometry invariant.

THEOREM 3.19. *Let G be a finitely generated group. Then G has a finite index nilpotent subgroup if and only if it has polynomial growth.*

CHAPTER 2

Negative curvature in spaces

There are many notions of curvature in spaces. To do geometric group theory, we are interested in formulating a notion that applies to metric spaces in general. This approach is informed by more classical notions of negative curvature in manifolds. Given a Riemannian manifold, the curvature may be formalised using *sectional curvature*: given two linearly independent vectors in a tangent space to a point, one calculates the Gaussian curvature of the surface with tangent plane equal to the span of these vectors.

Already, the topology of complete manifolds with everywhere non-positive sectional curvature is tied to group theory – the Cartan–Hadamard theorem tells us that the universal cover of such a manifold of dimension n is homeomorphic to $\mathbb{R}^n$. It follows that these manifolds are aspherical, and so their algebraic topology is largely determined by their fundamental groups. The most basic examples of manifolds of non-positive curvature are those with constant negative curvature: these are *hyperbolic manifolds*. The study of hyperbolic manifolds is incredibly vast and incredibly rich; we here give a brief overview of the basics in low dimensions, to give some intuition and motivation for the more abstract, metric, and combinatorial notions that will be the focus of most of this course.

1. Hyperbolic geometry

1.1. History. Hyperbolic geometry is the geometry of space with a constant negative curvature, and can be thought about in contrast to the geometry of space with zero curvature (i.e. Euclidean geometry) and constant positive curvature (i.e. spherical geometry). The development of hyperbolic geometry has a storied history, and was born out of an almost two-millennia-long attempt to reconcile a difficult tension in Euclid's classical axiomatisation of geometry.

The core point of contention was the nature of Euclid's fifth axiom, called the 'parallel postulate'. Contrary to the other four axioms (e.g. there exists a straight line between any two points, all right angles are equal), the parallel postulate is much more complicated, stating 'if two lines meet a third line, then the two lines will meet on the side of the third line for which the angle sum is less than the sum of two right angles'.

Coupled with the fact that more than half of the propositions in the first book of the Elements do not invoke the parallel postulate, it was widely believed that it should follow from the other axioms. Out of many attempts to prove this, it was gradually realised that the rejection of this axiom actually entails a consistent and robust geometry, often called 'absolute geometry', and that the truth of the parallel postulate in a particular model of geometry is independent of the other axioms. There are in fact only two models of absolute geometry, and they are exactly Euclidean and hyperbolic geometry, with the latter obtained by taking a negation of the parallel postulate instead of the postulate itself.

1.2. The Poincaré ball. We will write $\mathbb{H}^n$ for hyperbolic space of dimension n. This is, the unique simply connected Riemannian manifold of constant negative curvature -1. As a model for this space, we will take the open unit ball in Euclidean space $\mathbb{R}^n$, equipped with the metric

$$\mathrm{d}s^2 = \frac{4\|\mathrm{d}\mathbf{x}\|^2}{(1-\|\mathbf{x}\|^2)^2}.$$

This model is called the *Poincaré ball model* for hyperbolic space. There is no isometric embedding of $\mathbb{H}^n$ into Euclidean space of any dimension (unlike, say, a sphere with its intrinsic metric), so any such model must be far from distance-preserving. In fact, this model is *conformal* – it preserves angles – but it is easy to see that distances between points are heavily distorted from their Euclidean counterparts. Other common models are the hyperboloid model and the half space model; each comes with its own advantages and drawbacks.

Hyperbolic space has a natural bordification $\partial\mathbb{H}^n$, which we call the *space at infinity* or simply the *boundary*. From the ball model, this boundary is clear to see as the boundary sphere $\partial\mathbb{H}^n \cong S^{n-1}$. The geodesics in this space are given by diametrical lines and arcs of circles that are perpendicular to the boundary $\partial\mathbb{H}^n$. The group of isometries of $\mathbb{H}^n$ in this model is the Lie group $\mathrm{SO(n,1)}$ of special orthogonal matrices of signature $(n, 1)$. The space $\mathbb{H}^n$ is *homogenous* and *isotropic* – its group of isometries acts transitively on the space, and transitively on the tangent space at any given point.

Exercise 1.1. Verify using the path integral formula

$$\ell(\gamma) = \int_{t\in I} \frac{2|\gamma'(t)|}{1-\|\gamma(t)\|^2}\,\mathrm{d}t$$

for a path $\gamma\colon I \to \mathbb{H}^n$, that the geodesics in $\mathbb{H}^n$ are as described above.

Hyperbolic geometry has some interesting features that distinguish it from Euclidean geometry. The easiest to see of these is the non-parallelism of geodesics described in the previous section: this can be seen by the description of geodesics as above. In fact, geodesics will always diverge from each other rather quickly, in one direction or another. There is also the phenomenon of *ultra-parallelism*, where geodesics can share one endpoint in $\partial\mathbb{H}^n$. Such geodesics will stay a bounded distance from one another as they approach one point at infinity, and diverge in the other direction.

A second key feature of hyperbolic geometry is the uniform thinness of polyhedra. In Euclidean geometry, due to the existence of homothety, there are triangles of arbitrarily large area and arbitrarily large incircles. In hyperbolic geometry, this behaviour is forbidden. Let M be a compact Riemannian surface and recall the Gauss-Bonnet formula from differential geometry:

$$\iint_M K\mathrm{d}A + \int_{\partial M} k_g \mathrm{d}s = 2\pi\chi(M),$$

where K is the Gaussian curvature of M and k_g is the geodesic curvature of ∂M. With a smoothing argument, one can apply this formula to calculate the area of triangles. For an isometrically embedded triangle T in hyperbolic space, $K \equiv 1$, $\chi(T) = 1$, and contribution of the integral involving the boundary corresponds exactly to the sum of the external angles. Hence

$$\mathrm{area}(T) = \pi - (\alpha + \beta + \gamma),$$

where α, β, and γ are the internal angles of T. The area of a hyperbolic triangle is thus always bounded from above by π. It follows, for instance, that there is a uniform bound on the radius of an incircle in a hyperbolic triangle; they are all thin.

EXERCISE 1.2. Show explicitly that $\mathbb{H}^2$, and thus $\mathbb{H}^n$, has triangles that are uniformly thin in the above sense, with the constant $\frac{1}{2}\log 3$ as the bound on radii. (Hint: the worst you could do is an *ideal triangle*, one whose vertices lie on the boundary circle of $\mathbb{H}^2$.)

Another clear consequence of the above formula is the fact that the angle sum in a hyperbolic triangle is always *less* than π, contrasting the Euclidean case, where it is equal to π. In fact, the angle sum decreases proportionally to the area, with ideal triangles having the largest area and angle sum of zero.

A final key difference between Euclidean and hyperbolic geometry that we highlight here concerns balls. In Euclidean space of dimension n, a ball of radius r has volume proportional to r^n. Already, however, the area of a disc in the hyperbolic plane has grows exponentially with respect to the radius. In fact, the same is true even of the circumference of a circle in hyperbolic space. One may verify this by means of computing some not-too-complicated integrals. As a result, one sees that two different (unit-speed) geodesic rays in $\mathbb{H}^n$ with a shared origin 'diverge' exponentially quickly, in the sense that one must travel exponentially long distances with respect to a radius to get from a point on one to the other, outside a ball around the origin with that radius.

1.3. Surfaces and tessellations. In general, it is a fact that every hyperbolic manifold arises as the quotient $\mathbb{H}^n/\Gamma$ of hyperbolic space by a torsion-free discrete subgroup of isometries $\Gamma \leqslant \mathrm{Isom}(\mathbb{H}^n)$. This is easy to see if one assumes that $\mathbb{H}^n$ is the unique simply connected Riemannian manifold of constant curvature -1: each hyperbolic n-manifold M has universal cover $\widetilde{M} = \mathbb{H}^n$ (after possibly rescaling the metric) on which $\Gamma = \pi_1(M)$ acts by isometries. That Γ is torsion-free corresponds to the fact that the action of $\pi_1(M)$ is free and $M = \mathbb{H}^n/\Gamma$ has no singular points. This is a little abstract, so let us restrict our attention now to the two-dimensional case of the hyperbolic plane, where such a realisation can be explicitly computed and visualised.

Our construction will involve understanding polygons and tessellations of the hyperbolic plane. We sketch a proof of the following:

LEMMA 1.3. *Let $n, m \in \mathbb{N}$ be natural numbers with $\frac{1}{n} + \frac{1}{m} < \frac{1}{2}$. Then there is a tessellation of $\mathbb{H}^2$ by regular n-gons, with m different n-gons meeting at every vertex.*

PROOF. Very close to the origin in the Poincaré disc, the metric closely resembles that of Euclidean space. That is, there are regular n-gons centred on the origin, whose interior angle sum is arbitrarily close to $\frac{1}{2}(n-1)\pi$, the corresponding angle sum in Euclidean space. Following the above discussion on area, moving the vertices outward from the origin decreases this angle sum monotonically, and the sum approaches zero as the n-gon tends to an ideal n-gon. By a continuity argument, there are regular hyperbolic n-gons each of whose interior angles is equal to a given $0 < \theta < \frac{1}{2}(1 - \frac{1}{n})\pi$.

If θ in the above is taken of the form $\frac{2\pi}{m}$ for some natural number m, then we can obtain a tessellation of the hyperbolic plane by regular polygons, by reflecting such a polygon along its edges. The condition that $\theta = \frac{2\pi}{m} < \frac{1}{2}(1 - \frac{1}{n})\pi$ can be rearranged exactly into the hypothesis in the lemma, so it holds by assumption. $\square$

Using this, we may realise hyperbolic manifold structures on every surface that is not the torus or the sphere.

PROPOSITION 1.4. *For each $g \geq 2$, the surface Σ_g of genus g admits a Riemannian metric of constant negative sectional curvature. More precisely, there is a discrete torsion-free subgroup $\Gamma \leqslant \mathrm{Isom}(\mathbb{H}^2)$ such that Σ_g is isometric to $\mathbb{H}^2/\Gamma$.*

PROOF. Recall that Σ_g may be realised as a quotient of a (regular) $4g$-gon P. By Lemma 1.3, there is a tessellation of $\mathbb{H}^2$ by copies of P. Now $\mathrm{Isom}(\mathbb{H}^2)$ acts transitively on (oriented) line segments of the same length, so that there are isometries realising each of the side identifications appearing in the above quotient. Let Γ be the subgroup of $\mathrm{Isom}(\mathbb{H}^2)$ generated by these finitely many isometries.

As any element of Γ preserves the given tessellation of $\mathbb{H}^2$, it is straightforward to check that the action is properly discontinuous. Since $\mathbb{H}^2$ is a proper metric space, this means that Γ is a discrete subgroup of $\mathrm{Isom}(\mathbb{H}^2)$. Finally, observe that Γ fixes no points of $\mathbb{H}^2$. The existence of torsion in Γ would imply the existence of a fixed point, so Γ must be torsion-free. $\square$

The above is essentially a simple case of a more general theorem of Poincaré, which constructs discrete subgroups of $\mathrm{Isom}(\mathbb{H}^2)$ whose quotient realises any (orbi)surface with genus g and cone points of order $m_1, \ldots, m_n$, provided

$$2g - 2 + \sum_{i=1}^{n} \left(1 - \frac{1}{m_i}\right) \geq 0.$$

The quantity on the left is often called the *signature* of the surface. One can prove Poincaré's polygon theorem similarly to the above, but with a more involved construction of fundamental domain to account for the cone points.

It follows immediately from the above, together with the Milnor–Schwarz Lemma.

COROLLARY 1.5. *Let Σ be a surface of genus $g \geq 2$. Then $\pi_1\Sigma$ is quasi-isometric to $\mathbb{H}^2$.*

Discrete subgroups of $\mathrm{Isom}(\mathbb{H}^2)$ are often called *Fuchsian groups*, and discrete subgroups of $\mathrm{Isom}(\mathbb{H}^3)$ – and sometimes those of $\mathrm{Isom}(\mathbb{H}^n)$ – are called *Kleinian*.

1.4. The boundary. Any element of $\mathrm{Isom}(\mathbb{H}^n)$ has an induced action by homeomorphisms on the boundary $\partial\mathbb{H}^n$ of hyperbolic space, so there is a well-defined homomorphism $\mathrm{Isom}(\mathbb{H}^n) \to \mathrm{Homeo}(S^{n-1})$ for each $n \in \mathbb{N}$. As such, we can attempt to retrieve information about subgroups of isometries of $\mathbb{H}^n$ by analysing their action on the boundary. For individual isometries, this turns out to be very doable.

PROPOSITION 1.6. *Let $g \in \mathrm{Isom}(\mathbb{H}^n)$ be an isometry. Then either*

(1) g fixes a point in $\mathbb{H}^n$;

(2) g fixes no points in $\mathbb{H}^n$ and exactly one point in $\partial\mathbb{H}^n$; or

(3) g fixes no points in $\mathbb{H}^n$ and exactly two points in $\partial\mathbb{H}^n$.

We call the isometries *elliptic*, *parabolic*, and *loxodromic* respectively in the above cases. In each of the above cases, the geometry of the isometry g can be effectively described. If g is elliptic, then it is a rotation around its fixed point in $\mathbb{H}^n$, since the stabiliser of any point in $\mathbb{H}^n$ is the Lie group $O(n)$. If g is parabolic, then it fixes any *horosphere* centred around its fixed point in $\partial\mathbb{H}^n$. A horosphere is the limit of a sequence of spheres of increasing radii with a shared point of tangency, and its centre is the point in the boundary that meets the diameter of these spheres. A horosphere is a copy of the Euclidean plane, embedded in $\mathbb{H}^n$ with exponential distortion of the metric. Lastly, a loxodromic isometry fixes a bi-infinite geodesic joining its two fixed points in $\partial\mathbb{H}^n$, and acts as a translation when restricted to this axis.

For more general subgroups of isometries than cyclic ones, the situation is naturally more complicated. Here, the action of the subgroup on a particular subset of the boundary known as its *limit set* becomes important.

DEFINITION 1.7. Let $G \leqslant \mathrm{Isom}(\mathbb{H}^n)$. The *limit set* of G is the subset $\Lambda G \subseteq \partial\mathbb{H}^n$ of accumulation points of G-orbits in $\mathbb{H}^n$.

EXERCISE 1.8. Show that ΛG is the smallest G-invariant closed subset of $\partial\mathbb{H}^n$, and that ΛG is a perfect compactum unless G has a finite index cyclic subgroup.

The study of limit sets and their geometric properties is of much interest. They are in general fractal subsets of the boundary. Various facts about a subgroup can be determined from its limit set; we do not pursue these here, but will return to the topic when discussing boundaries of abstract hyperbolic groups later.

2. Hyperbolic metric spaces

We now introduce a notion of negative curvature for metric spaces. Our definition will be modelled on a key property of the classical hyperbolic spaces of the previous section: it will state that every geodesic triangle is uniformly thin. That is, triangles in these spaces will look somewhat like tripods. Triangles are the most basic shapes in a geodesic space, and so, as we will see, this assumption has some strong consequences for the geometry of these spaces and the groups that act on them.

We will need some preliminary definitions.

DEFINITION 2.1 (Gromov product). Let (X, d) be a metric space, and $x, y, z \in X$ be points. The *Gromov product* of x and y with respect to z is

$$\langle x, y\rangle_z = \frac{1}{2}\Big(\mathrm{d}(x,z) + \mathrm{d}(y,z) - \mathrm{d}(x,y)\Big).$$

One can think of the Gromov product $\langle x, y\rangle_z$ as an abstracted notion of the 'angle' spanned by x and y with respect to z. Indeed, in Euclidean space, this Gromov product is exactly the distance of the point z to the points on $[x, z]$ and $[y, z]$ that touch the incircle of the triangle with vertices x, y, and z – up to homothety, this is determined by the angle these two lines make.

DEFINITION 2.2 (Thin triangles). Let Δ be a geodesic triangle with vertices x, y, and z in a metric space X, and let $\delta \geq 0$. Call T_Δ the tripod with leg lengths $\langle x, y\rangle_z, \langle x, z\rangle_y$, and $\langle y, z\rangle_z$. There is a unique map $\varphi\colon \Delta \to T_\Delta$ such that x, y, and z map to the extremal vertices of T_Δ and φ restricts to an isometry on each side of Δ. We say Δ is *δ-thin* if $\operatorname{diam} \varphi^{-1}(\{t\}) \leq \delta$ for all $t \in T_\Delta$.

Definition 2.3 (Hyperbolic metric space). Let X be a geodesic metric space. If there is $\delta \geq 0$ such that every geodesic triangle in X is δ-thin, we say that X is a *δ-hyperbolic metric space*. We simply call X a *hyperbolic metric space* if there is some $\delta \geq 0$ such that it is a δ-hyperbolic metric space.

Exercise 2.4. Show that a geodesic space is 0-hyperbolic if and only if it is an *$\mathbb{R}$-tree*: a space in which every pair of points is connected by a unique arc.

Example 2.5. We saw in the previous section that $\mathbb{H}^n$ is δ-hyperbolic with hyperbolicity constant $\delta = \frac{1}{2}\log 3$.

Example 2.6. The plane $\mathbb{R}^2$ is not a hyperbolic metric space, as for any $\delta \geq 0$, any equilateral triangle with side lengths greater than 2δ is not δ-thin.

There are many alternative formulations of the thin triangles condition. One that is very commonly used and can be useful is the *slim triangles* formulation.

Definition 2.7. Let Δ be a geodesic triangle in metric space X, and let $\delta \geq 0$. We say that Δ is *δ-slim* if each side of Δ is contained in a δ-neighbourhood of the union of the other two sides.

Of course, using slim triangles instead of thin triangles gives an identical characterisation of hyperbolic metric spaces, up to a small change in the constant in the definitions.

Exercise 2.8. Show that every δ-thin triangle is δ-slim, and that every δ-slim triangle is also 2δ-thin.

Exercise 2.9. Show that a if geodesic space X is δ-hyperbolic then it satisfies the *four-point condition*: for all $x, y, z, w \in X$, we have

$$\mathrm{d}(x, y) + \mathrm{d}(z, w) \leq \max\{\mathrm{d}(x, z) + \mathrm{d}(y, w), \mathrm{d}(x, w) + \mathrm{d}(y, z)\} + 2\delta.$$

Further, show that if X satisfies the four-point condition for some $\delta \geq 0$, there is $\delta' \geq 0$ such that X is δ'-hyperbolic.

Remark 2.10. The four-point condition above can of course be formulated for any metric space, without any assumption on whether or not geodesics exist. This is frequently useful, as it allows us to talk about hyperbolicity of discrete metric spaces, such as groups equipped with a word metric.

An important feature of hyperbolic metric spaces is the 'stability' of quasi-geodesics: all quasi-geodesics actually follow geodesics between their endpoints uniformly closely. This fact is usually referred to as the *Morse Lemma*. Note that this feature is particular to hyperbolic metric spaces; the following example shows that it can dramatically fail outside of this setting.

Example 2.11. Consider the Cayley graph of $\mathbb{Z}^2$ with the standard generating set. Then take the concatenation of three geodesics of length n, one vertical geodesic going up, one horizontal going right, and one vertical going down, is a $(3, 0)$-quasi-geodesic.

This path contains points that are a distance of n from the unique geodesic joining its endpoints (which is a horizontal path of length n), and we may take n to be arbitrarily large.

In order to obtain the stability statement, we must first obtain an estimate on the length of paths different from geodesics. Simply stated, we have that the length of a path grows at least exponentially with its distance from a geodesic between its endpoints.

LEMMA 2.12. *Let X be a δ-hyperbolic space, and $x, y \in X$. If $\gamma\colon I \to X$ is a continuous rectifiable path between x and y, then*

$$d(z, \gamma(I)) \leq \delta \max\{0, \log_2 \ell(\gamma)\} + 2$$

for any z on a geodesic between x and y.

PROOF. Let $n = \lceil \log_2 \ell(\gamma) \rceil$ and suppose that γ is a parameterisation proportional to arc-length, so that we may write $I = [0, 1]$. If $\ell(\gamma) \leq 1$ then there is nothing to prove, so suppose otherwise. It follows that $n \geq 1$ is a positive natural number.

Let z be a point on a geodesic from $x = \gamma(0)$ to $y = \gamma(1)$. Since triangles in X are δ-slim, for any $k \geq 0$ and $i = 1, \ldots, 2^k$, any point on a geodesic $[\gamma(\frac{i-1}{2^k}), \gamma(\frac{i}{2^k})]$ is at most distance δ from a geodesic of the form $[\gamma(\frac{j-1}{2^{k+1}}), \gamma(\frac{j}{2^{k+1}})]$ for some $1 \leq j \leq 2^{k+1}$. It follows by a finite induction that for any $k \geq 0$, there is $1 \leq i \leq 2^k$ such that

$$d(z, p) \leq \delta k. \tag{2.1}$$

where p is a geodesic of the form $[\gamma(\frac{i-1}{2^k}), \gamma(\frac{i}{2^k})]$.

Now for any $i = 1, \ldots, 2^n$, the length of the subpath γ restricted to $[\frac{i-1}{2^n}, \frac{i}{2^n}]$ is at most 1 by the choice of n. Hence any point on a geodesic between $\gamma(\frac{i-1}{2^n})$ and $\gamma(\frac{i}{2^n})$ is a distance of less than 1 from $\gamma(I)$. Combined with (2.1) applied to the case $k = n$, this fact gives the required inequality. □

PROPOSITION 2.13 (Morse Lemma). *Let X be a δ-hyperbolic space, $\lambda \geq 1$, and $c \geq 0$. There is a constant $M = M(\lambda, c, \delta) \geq 0$ such that any (λ, c)-quasi-geodesic segment in X is a Hausdorff distance of at most M from any geodesic segment between its endpoints.*

PROOF. Let $\gamma\colon I \to X$ be a (λ, c)-quasi-geodesic segment, and write x and y for its endpoints. Applying the lemma on continuous quasi-geodesics, there is a (λ, c')-quasi-geodesic $\gamma'\colon J \to X$ with the same endpoints as γ, lying a Hausdorff distance of at most c' from γ. We may suppose that γ' is parametrised by arc-length. Fix a geodesic p whose endpoints are also x and y. Let z be a point on p maximising $r = d(z, \gamma')$, which exists by continuity. We will exhibit an absolute bound on r.

Let x' and y' be points on p between x and z and y and z respectively, with $d(x', z) = d(y', z) = 2r$ (choosing $x' = x$ or $y' = y$ if $d(x, z) \leq 2r$ or $d(y, z) \leq 2r$ respectively). By the definition of z, we have

$$d(x', \gamma') \leq r \quad \text{and} \quad d(y', \gamma') \leq r.$$

Let $s, t \in [a, b]$ be points such that $d(x', \gamma'(s))$ and $d(y', \gamma'(t))$ realise the distances from x' and y' to γ', which again exist by continuity. We will say $s \leq t$, swapping the names if otherwise. It follows then, that

$$d(\gamma'(s), \gamma'(t)) \leq d(\gamma'(s), x') + d(x', y') + d(y', \gamma'(t)) \leq 6r$$

Write ξ for concatenation of the subpath of γ' between $\gamma'(s)$ and $\gamma'(t)$ with geodesics $[x', \gamma'(s)]$ and $[\gamma'(t), y']$. Now using the fact that γ' is a quasi-geodesic, the above implies that

$$\ell(\xi) \leq r + \ell(\gamma'|_{[s,t]}) + r \leq (6\lambda + 2)r + \lambda c'.$$

Then by Lemma 2.12 and the choice of z, we have

$$r = \mathrm{d}(z, \xi) \leq \delta \log_2((6\lambda + 2)r + \lambda c') + 2.$$

Rearranging slightly, we have

$$2^{\frac{1}{\delta}(r-1)} \leq (6\lambda + 2)r + \lambda c'$$

so that an exponential function in r is bounded above by a linear function in r. This imposes a uniform bound M' on r depending on these two functions, which in turn depend only on δ, λ, and c. Thus p is contained in a M'-neighbourhood of γ'.

We have thus shown that $p \subseteq N_{M'}(\gamma(I))$. To show the converse inclusion, let $t \in J$. Of course, if t is an endpoint of J we are done, so suppose otherwise. By continuity of γ', there is a point z of p and numbers $s < t < s'$ such that $\mathrm{d}(\gamma'(s), z) \leq M'$ and $\mathrm{d}(\gamma'(s'), z) \leq M'$. Thus $\mathrm{d}(\gamma'(s), \gamma'(s')) \leq 2M'$. As γ' is a quasi-geodesic, this implies

$$\ell(\gamma'|_{[s,s']}) \leq 2\lambda M' + \lambda c'.$$

Whence $\mathrm{d}(\gamma(t), p) \leq (2\lambda + 1)M' + \lambda c'$. Finally, setting $M = (2\lambda + 1)M' + (\lambda + 1)c'$ completes the proof. $\square$

Note that the key point in the proof was the exponential length estimate on non-geodesic paths in hyperbolic metric spaces. The condition of quasigeodesicity gives a linear upper bound to this exponential bound, but the proof works just as well if the distortion on the length of the path is subexponential.

The most important consequence of the Morse Lemma is the quasi-isometry invariance of hyperbolicity.

THEOREM 2.14. *For any $\lambda \geq 1, c \geq 0$, and $\delta \geq 0$ there is a constant $\delta' \geq 0$ such that the following is true.*

Let X and Y be geodesic spaces, and suppose that X is a δ-hyperbolic metric space. If $f \colon Y \to X$ is a (λ, c)-quasi-isometric embedding, then Y is a δ'-hyperbolic metric space.

PROOF. Let Δ be a geodesic triangle in Y with sides p, q, and r. The paths $f(p), f(q)$, and $f(r)$ are (λ, c)-quasi-geodesics in X. Since X is δ-hyperbolic, the Morse Lemma gives us a constant $M = M(\lambda, c, \delta) \geq 0$ such that each of these paths is within an M-Hausdorff neighbourhood of any geodesic between their endpoints.

As geodesic triangles are δ-slim in X, any side of a geodesic triangle whose vertices are the endpoints of $f(p), f(q)$, and $f(r)$ is contained in a δ-neighbourhood of the other two. Thus $f(p)$ is contained in a $(\delta + 2M)$-neighbourhood of $f(q) \cup f(r)$. Let x be a point in p. By the above there is a point y in q or r such that $\mathrm{d}_X(f(x), f(y)) \leq \delta + 2M$. Now as f is a (λ, c)-quasi-isometry, we have

$$\mathrm{d}_Y(x, y) \leq \lambda\, \mathrm{d}_X(f(x), f(y)) + \lambda c \leq (\delta + 2M + c)\lambda.$$

By symmetry, analogous inequalities are true for points on q and on r. Hence every geodesic triangle in Y is $(\delta + 2M + c)\lambda$-slim. We see that Y is δ'-hyperbolic, where $\delta' = 2(\delta + 2M + c)\lambda$. $\square$

Corollary 2.15. *Let X and Y be quasi-isometric geodesic spaces. Then X is a hyperbolic metric space if and only if Y is a hyperbolic metric space.*

Note that δ-hyperbolicity is not invariant under quasi-isometry – the constant of hyperbolicity may change depending on the quasi-isometry constants. Indeed, given a δ-hyperbolic metric space X, one can attach a sphere of diameter R to obtain a space X' that is not δ-hyperbolic if $R > \delta$. Of course, X and X' are $(1, \pi R)$-quasi-isometric, and X' is clearly δ'-hyperbolic, where $\delta' = \delta + R$.

One of the central features of hyperbolicity is the prevalence of many 'local-to-global' phenomena: results that conclude something about the large-scale, global geometry of a space from small-scale, local conditions. Frequently useful is the sufficient condition for quasi-geodesics below.

Definition 2.16 (Local quasi-geodesic). Let X be a metric space, $\lambda \geq 1, c \geq 0$, and $k \geq 0$. A rectifiable path $p\colon I \to X$ is a *k-local (λ, c)-quasi-geodesic* if each subpath q of p with $\ell(q) \leq k$ is a (λ, c)-quasi-geodesic.

Theorem 2.17. *Let X be a δ-hyperbolic space, $\lambda \geq 1$, and $c \geq 0$. There are $\lambda' \geq 1, c' \geq 0$, and $k \geq 0$ such that every k-local (λ, c)-quasigeodesic is (λ', c')-quasigeodesic.*

Proof. Let $M = M(\lambda, c, \delta)$ be the constant of the Morse Lemma, and let $k = 2\lambda(2M + 4\delta + c + 1)$. Let $\gamma\colon I \to X$ be a k-local (λ, c)-quasigeodesic, and let $p\colon J \to X$ be a geodesic with the same endpoints. Suppose both γ and p are parametrised by arc-length. Increasing c, we may assume γ is a continuous path, so that the upper bound $\mathrm{d}(\gamma(t), \gamma(t')) \leq |t - t'|$ holds trivially for any $t, t' \in I$. The idea is that points at $k/2$-intervals along γ will project to points that make uniform progress along on p.

We first claim that $\gamma(I)$ is contained in a uniform neighbourhood of $p(J)$. Let $t \in I$ maximise $\mathrm{d}(\gamma(t), p(J))$, and let $s, u \in I$ be the points $s = \max\{t - \frac{k}{2}, \inf(I)\}$ and $u = \min\{t + \frac{k}{2}, \sup(I)\}$. Let $s', u' \in J$ be such that $p(s')$ and $p(u')$ are the closest points on p to $\gamma(s)$ and $\gamma(u)$. By the Morse Lemma, $\gamma(t)$ is M-close to a point x on a geodesic $[\gamma(s), \gamma(u)]$. Consider the rectangle whose vertices are $\gamma(s), \gamma(u), p(s')$, and $p(u')$. By hyperbolicity, this rectangle is 2δ-slim, so x is 2δ-close to a point w on $[\gamma(s), p(s')], [\gamma(u), p(u')]$, or $[p(s'), p(u')]$. We will rule out the former two possibilities, which completes the claim.

Indeed, suppose that w is a point on $[\gamma(s), p(s')]$ with $\mathrm{d}(x, w) \leq 2\delta$. Thus we have $\mathrm{d}(\gamma(t), w) \leq 2\delta + M$. Since γ is a (λ, c)-quasigeodesic when restricted to $[s, t]$, we have $\mathrm{d}(\gamma(s), \gamma(t)) \geq \frac{1}{2\lambda}k - c$. The triangle inequality then gives us

$$-\,\mathrm{d}(\gamma(s), w) \leq 2\delta + M - \frac{1}{2\lambda}k + c. \tag{2.2}$$

Now using the fact that $[\gamma(s), p(s')]$ is a geodesic on which w lies,

$$\begin{aligned}\mathrm{d}(\gamma(t), p(s')) &\leq \mathrm{d}(\gamma(t), w) + \mathrm{d}(w, p(s')) \\ &\leq 2\delta + M + \mathrm{d}(\gamma(s), p(s')) - \mathrm{d}(\gamma(s), w).\end{aligned}$$

Whereby (2.2) and the choice of k allows us to conclude

$$\mathrm{d}(\gamma(t), p(s')) \leq \mathrm{d}(\gamma(s), p(s')) + 4\delta + 2M + c - \frac{1}{2\lambda}k < \mathrm{d}(\gamma(s), p(s')),$$

which contradicts the choice of t. A symmetrical argument applies to show that w does not lie on $[\gamma(u), p(u')]$. Therefore $\mathrm{d}(\gamma(t), p(J)) \leq M + 2\delta$.

We are now ready to prove the main statement. Let $t_0 < \cdots < t_n$ be a partition of I such that $t_i - t_{i-1} = k/2$ for $i = 1, \ldots, n-1$, and $t_n - t_{n-1} \leq k/2$. For each $i = 0, \ldots, n$, let $s_i \in J$ be a point with $\mathrm{d}(\gamma(t_i), p(s_i)) \leq 2\delta + M$, as guaranteed to exist by the claim above. By techniques similar to the proof of the claim, one can show $s_{i-1} < s_i$ for all $i = 1, \ldots, n$. Moreover, by local quasigeodesicity of γ and the choice of k,

$$\mathrm{d}(p(s_{i-1}), p(s)) \geq \frac{1}{2\lambda}k - c - 2M - 4\delta \geq 1,$$

for any $i = 1, \ldots, n$. As $s_0 < \cdots < s_n$, it follows that for any $0 \leq i < j \leq n$

$$\mathrm{d}(p(s_i), p(s_j)) \geq j - i.$$

Let $t, t' \in I$ be arbitrary. There are $i, j \in \mathbb{N}$ minimising $|t - t_i|$ and $|t' - t_j|$. By construction, these quantities are at most $\frac{1}{4}k$, and $|t_i - t_j| = \frac{1}{2}|i - j|k$ (out of laziness, we ignore the edge case that i or j is equal to n here). Combining all of the above

$$\begin{aligned}
\mathrm{d}(\gamma(t), \gamma(t')) &\geq \mathrm{d}(\gamma(t_i), \gamma(t_j)) - \mathrm{d}(\gamma(t), \gamma(t_i)) - \mathrm{d}(\gamma(t'), \gamma(t_j)) \\
&\geq \mathrm{d}(p(s_i), p(s_j)) - 2M - 4\delta - \frac{1}{2}\lambda k - 2c \\
&\geq |i - j| - 2M - 4\delta - \frac{1}{2}\lambda k - 2c \\
&\geq \frac{2}{k}|t_i - t_j| - c' \\
&\geq \frac{1}{\lambda'}|t - t'| - c',
\end{aligned}$$

where $\lambda' = \frac{1}{2}k$ and $c' = 2M + 4\delta + 2c + \frac{1}{2}\lambda k$. □

There are many refinements one can apply to the above statement. For instance, the multiplicative constant λ' can be made arbitrarily close to λ, provided one insists that the local quasigeodesics are quasigeodesic on a sufficiently large scale.

Remark 2.18. Interestingly, the converse of the above theorem also holds: if a geodesic space has the property that all (large enough scale) local quasigeodesics are in fact quasigeodesics, the space is hyperbolic. Thus such local-to-global properties are not only characteristic of hyperbolicity, but unique to it. Indeed, this is true for a number of the basic theorems one proves about hyperbolicity. For instance, a geodesic space where the Morse Lemma holds is also necessarily a hyperbolic metric space.

We state without proof a useful and philosophically important result about hyperbolic spaces. It says that any finite set of points in a hyperbolic space can be approximated uniformly well by a tree having those points as vertices. This adds to the intuition that hyperbolic spaces are really like thickened trees.

Theorem 2.19. *Let X be a δ-hyperbolic metric space. There is a function $h\colon \mathbb{N} \to [0, \infty)$ such that if $x_1, \ldots, x_n \in X$ are points, there is an embedded simplicial tree $T \subseteq X$ with $x_1, \ldots, x_n$ as vertices with*

$$\mathrm{d}_T(x_i, x_j) \leq \mathrm{d}_X(x_i, x_j) + \delta h(n)$$

for any $i, j = 1, \dots, n$. *Moreover,* $h(n) = O(\log n)$.

3. Quasiconvex subspaces

In the study of geodesic metric spaces generally, the most well-behaved subspaces are the *convex subspaces*, subspaces that contain every geodesic with endpoints in the subspace. For a convex subspace, the intrinsic length metric of the subspace naturally coincides with the induced metric it inherits from the ambient space, so that the geometry of the subspace respects the geometry of the space it lives in. Since we are interested in coarse geometric properties of metric spaces, we will require a coarse version of this notion.

DEFINITION 3.1 (Quasiconvex subset). Let X be a geodesic space and $\sigma \geq 0$. We say that a subspace $Y \subseteq X$ is *σ-quasiconvex* if the image of any geodesic in X with endpoints in Y is contained in a σ-neighbourhood of Y.

It is a general fact that quasiconvex subsets are quasi-isometrically embedded in their ambient spaces. This mirrors the fact that convex subspaces are isometrically embedded in geodesic spaces. We will need some notation to be able to state this precisely.

DEFINITION 3.2 (Coarse metric). Let (X, d_X) be a metric space, $r \geq 0$. The the metric $\mathrm{d}_{X,r}$ on X is defined to $\mathrm{d}_{X,r}(x, y) = n$, where n is the minimal integer such that there is a sequence of points $x_0, \dots, x_n \in X$ with $x_0 = x, x_n = y$, and $\mathrm{d}_X(x_{i-1}, x_i) \leq r$ for all $i = 1, \dots, n$.

REMARK 3.3. This is the same as the edge path metric on the 1-skeleton of the Rips complex $P_r(X)$, which we will define in Section 7 of Chapter 3.

LEMMA 3.4. *Let (X, d_X) be a geodesic space, and $Y \subseteq X$ a subspace with induced metric $\mathrm{d}_Y = \mathrm{d}_X|_{Y\times Y}$. If there is $\sigma \geq 0$ such that Y is σ-quasiconvex, then the inclusion map $(Y, \mathrm{d}_{Y,r}) \to (X, \mathrm{d}_X)$ is a $(r, 1)$-quasi-isometry for any $r \geq 2\sigma + 1$.*

PROOF. Let $a, b \in Y$ be points and $r \geq 2\sigma + 1$,. Let $p\colon I \to X$ be a geodesic in X with endpoints a and b. Take a partition $t_0 < \dots < t_n$ of I where $t_i - t_{i-1} = 1$ for $i = 1, \dots, n-1$ and $t_n - t_{n-1} \leq 1$. We have that $n - 1 \leq \mathrm{d}_X(a, b) \leq n$. For each $i = 0, \dots, n$, there is $y_i \in Y$ with $\mathrm{d}_X(y_i, p(t_i)) \leq \sigma$. Of course, we can take $y_0 = a$ and $y_n = b$. Thus for each $i = 1, \dots, n$, we have

$$\mathrm{d}_Y(y_{i-1}, y_i) \leq 2\sigma + 1 \leq r.$$

It follows that

$$\mathrm{d}_{Y,r}(a, b) \leq n \leq \mathrm{d}_X(a, b) + 1.$$

The inequality $\mathrm{d}_X(a, b) \leq r\,\mathrm{d}_{Y,r}(a, b)$ is immediate, as every path in Y yields a path in X with length multiplied by at most r. Hence the inclusion map $Y \to X$ is a $(r, 1)$-quasi-isometric embedding. □

LEMMA 3.5. *Let X be a hyperbolic metric space, $Y \subseteq X$ a geodesic subspace. If Y is quasiconvex, then it is a hyperbolic metric space.*

PROOF. This is an immediate consequence of Lemma 3.4 and Theorem 2.14. □

In hyperbolic spaces, the converse of Lemma 3.4 also holds true.

LEMMA 3.6. *Let X be a δ-hyperbolic metric space, $\lambda \geq 1$ and $c \geq 0$. If Y is a geodesic space and $f\colon Y \to X$ is a quasi-isometric embedding, then there is $\sigma = \sigma(\lambda, c, \delta) \geq 0$ such that $f(Y)$ is σ-quasiconvex.*

PROOF. Let $M = M(\lambda, c, \delta)$ be the constant of the Morse Lemma. Any two points in Y may be joined by a geodesic, so that any two points in $f(Y)$ may be joined by a (λ, c)-quasigeodesic lying entirely in Y. The Morse Lemma implies that any geodesic in X with the same endpoints as such a quasi-geodesic is contained in an M-neighbourhood of it. Hence $f(Y)$ is σ-quasiconvex with $\sigma = M$. $\square$

4. The boundary of a hyperbolic metric space

In this ssection we introduce a natural bordification of hyperbolic metric spaces. Intuitively, this is the space of 'endpoints' of geodesic rays in the space. The study of boundaries of hyperbolic metric spaces is not only useful – as we shall later see – to the theory of groups acting on hyperbolic spaces, but also plays an essential role in the understanding the metric geometry of these spaces.

DEFINITION 4.1 (Gromov boundary). Let X be a hyperbolic metric space. We say two geodesic rays are *asymptotic* if their images are a finite Hausdorff distance apart. The *Gromov boundary* of X is the set ∂X of equivalence classes of geodesic rays in X, up to the relation of being asymptotic.

EXERCISE 4.2. Suppose that X is a proper hyperbolic metric space. Show that for any $x \in X$ and $p \in \partial X$, there is a geodesic ray based at x whose endpoint is p. Show that for any $p, q \in \partial X$, there is a bi-infinite geodesic line whose endpoints are p and q.

(Hint: Construct a sequence of finite approximating geodesic segments, then apply the Arzelà–Ascoli theorem and properness to conclude.)

It follows from the above exercise that for a proper hyperbolic space X, the set ∂X is in bijection with the classes or rays based at any particular given point $x \in X$. We will find it convenient to formulate some things regarding the boundary using this, though we must check independence from the choice of basepoint whenever we do so.

Many constructions and statements we present here will hold true for general hyperbolic spaces, but it turns out that if one does not assume that the space is proper, then many additional technicalities arise. For instance, in the above exercise, one would need to replace 'geodesic' with '$(1, 20\delta)$-quasigeodesic'. Since we will only ever be working with proper spaces in practice, we will usually include this assumption.

The following exercise demonstrates one such complication that comes with not assuming properness, even in the simple setting of trees.

EXERCISE 4.3. Show that if X is an unbounded proper hyperbolic space, then ∂X is non-empty. To contrast, construct an unbounded tree T for which ∂T is empty.

The Gromov boundary carries a natural topology, wherein we declare that two points are close if they have representative geodesics that stay close for a long time.

DEFINITION 4.4 (Topology on the boundary). Let X be a proper hyperbolic metric space and fix a point $x \in X$. For $p \in \partial X$ and $r > 0$, define the set

$$U(p, r) = \left\{ q \in \partial X \;\middle|\; \begin{array}{c} \gamma, \xi\colon [0, \infty) \to X \text{ are geodesic rays with } \gamma(0) = \xi(0) = x, \\ p = [\gamma], q = [\xi], \text{ and } \liminf_{t \to \infty} \langle \gamma(t), \xi(t) \rangle_x \geq r \end{array} \right\}.$$

The sets $U(p,r)$ form a basis of neighbourhoods of p for a topology on ∂X.

Exercise 4.5. Verify that the collection of sets $U(p,r)$ actually define a neighbourhood basis for a topology on ∂X. That is, if $p, p' \in \partial X$ and $r, r' \geq 0$, then there is $q \in \partial X$ and $s \geq 0$ such that

$$U(q,s) \subseteq U(p,r) \cap U(p',r').$$

Further, show that this topology is independent of the choice of basepoint.

Example 4.6.

- If T is a tree with valence at least 3 at every vertex, then ∂T is a Cantor set.
- The boundary of $\mathbb{H}^n$ is the sphere S^{n-1}.

Exercise 4.7. Pick some basepoint x in a proper hyperbolic metric space X. We say a sequence $(x_i)_{i\in\mathbb{N}}$ *converges at infinity* if $\liminf_{i,j\to\infty}\langle x_i, x_j\rangle_x = \infty$. Two such sequences $(x_i)_{i\in\mathbb{N}}$ and $(y_i)_{i\in\mathbb{N}}$ are said to be *asymptotic* if $\liminf_{i,j\to\infty}\langle x_i, y_j\rangle_x = \infty$. Show that the set of equivalence classes of asymptotic sequences is in bijection with ∂X. For a sequence (x_i) converging at infinity, we write that $x_i \to q \in \partial X$ if q is the image of the class (x_i) represents under this bijection.

Further, we equip this set with a topology generated by the basis of neighbourhoods

$$V(p,r) = \left\{ q \in \partial X \;\middle|\; \begin{array}{l} (x_i)_{i\in\mathbb{N}}, (y_i)_{i\in\mathbb{N}} \text{ are sequences converging at infinity} \\ \text{with } x_i \to p, y_i \to q, \text{ and } \liminf_{i,j\to\infty}\langle x_i, y_j\rangle_x \geq r \end{array} \right\}.$$

Show that this topology coincides with that of Definition 4.4.

Definition 4.8 (Gromov product at infinity). Let X be a proper hyperbolic metric space and $x \in X$. Given $p, q \in X \cup \partial X$, we write

$$\langle p, q\rangle_x = \sup \liminf_{i,j\to\infty} \langle x_i, y_j\rangle_x$$

where the supremum is taken over all sequences (x_i) and (y_j) in X converging to p and to q respectively.

Exercise 4.9. Let X be a proper δ-hyperbolic metric space, $x \in X$ and $p, q \in X \cup \partial X$. Show that if $\langle p, q\rangle_x \leq r$, there is a point z on any geodesic line $[p, q]$ with $\mathrm{d}(x, z) \leq r + 2\delta$.

Boundaries of hyperbolic spaces admit metrics the induce the natural topology above. In particular, they are metrisable spaces, and so inherit all the wonderful properties of metric spaces (for example, they are regular, paracompact, and so on).

Definition 4.10 (Visual metric). Let X be a proper hyperbolic metric space and let $a > 1$. Pick a basepoint $x \in X$ and for any $p, q \in \partial X$, choose a bi-infinite geodesic in X joining p and q. Call $l_x(p,q)$ the distance of the geodesic to the point x.

A metric d on ∂X is called a *visual metric with parameter a* if there is a constant C such that

$$\frac{1}{C} a^{-l_x(p,q)} \leq \mathrm{d}(p,q) \leq C a^{-l_x(p,q)}.$$

It is not very hard to construct explicit visual metrics on boundaries of hyperbolic spaces, but we will their existence for granted here. Note that visual metrics are not canonical: there may be many depending on the choice of parameter. However, all visual metrics on a proper hyperbolic metric space are *quasi-symmetric*. That is to say, there is a self-homeomorphism of the space mapping one metric to the other, which preserves annuli in the space in a uniform way. A *quasi-conformal map* is a map that is quasi-symmetric and has quasi-symmetric inverse. The metric structure of the boundary up to quasi-conformal transformations is a canonical invariant of the space. We will not give precise definitions or pursue these notions further here.

PROPOSITION 4.11. *Let X be a proper hyperbolic metric space. Then ∂X is compact.*

PROOF. Observe that the basis of neighbourhoods $U(p,r)$ define the same topology as in Definition 4.4 if one takes $r > 0$ ranging over the rational numbers. Thus ∂X is first countable. Since ∂X is metrisable, this means that compactness is equivalent to sequential compactness.

Sequential compactness is a quick consequence of properness and the Arzelà–Ascoli theorem. We may take a sequence of points (p_i) in ∂X and geodesic rays (γ_i) representing these points, and the aforementioned theorem tells us that some subsequence (γ_{n_i}) of these converges to a ray γ. It follows that $a = [\gamma]$ is a limit of the subsequence (p_{n_i}). □

One can extend the topology in Definition 4.4 to include points within the space, by allowing the both geodesic rays and segments in the definition. This gives a topology on $X \cup \partial X$ for which the subspace topology on X agrees with the topology induced by the metric on X. Indeed, essentially the same argument as the previous lemma shows that $X \cup \partial X$ is compact with this topology, so that ∂X can be thought of as a compactification of X.

PROPOSITION 4.12. *Let X be a proper hyperbolic space. Then $X \cup \partial X$ is compact.*

There is in fact the maximal possible diversity among the spaces that can be realised as boundaries of hyperbolic spaces.

PROPOSITION 4.13. *For any compact metrisable space M, there is a proper hyperbolic metric space X with $\partial X \cong M$.*

PROOF. Every regular second countable Hausdorff space is homeomorphic to a subspace of the Hilbert cube $C = \prod_{n\in\mathbb{N}}[0, \frac{1}{n}]$ by the Urysohn metrisation theorem. Certainly, M satisfies this criterion as it is a compact metric space. Further, C can be embedded into the unit sphere of a separable Banach space: C already naturally lies in the unit ball of a separable Banach space $\mathcal{H}'$, so project it to the upper hemisphere of the unit ball in $\mathcal{H} = \mathcal{H}' \times \mathbb{R}$. Thus there is a topological embedding $\iota\colon M \to \mathcal{S}$, where $\mathcal{S}$ is the unit sphere of a separable Banach space $\mathcal{H}$.

We equip the open unit ball $\mathcal{B}$ with with the *Cayley-Klein metric* d, which is defined as follows. For any $x, y \in \mathcal{B}$, let $p, q \in \mathcal{S}$ be the points of $\mathcal{S}$ meeting the line $L = \{\lambda p + (1-\lambda)q \mid \lambda \in \mathbb{R}\}$. The metric is defined as

$$\mathrm{d}(x,y) = \frac{1}{2}\left|\log\frac{\|x-q\|\|y-p\|}{\|x-p\|\|y-q\|}\right|.$$

Note that the quantity in the logarithm is the *cross-ratio*, which may be familiar from projective geometry, and plays a large role in hyperbolic geometry. This metric makes $\mathcal{B}$ a model for an infinite dimensional hyperbolic space, and $\mathcal{S}$ is its ideal boundary.

One can verify that the geodesics with respect to this metric are exactly intersections of lines (that is, one-dimensional affine linear subspaces) of $\mathcal{H}$ with $\mathcal{B}$. It follows that the convex hull of a subset of a subset of $\mathcal{B}$ with respect to the metric d is exactly the convex hull with respect to $\mathcal{H}$, the set of all linear combinations of points in the subset.

Let $X \subseteq \mathcal{B}$ be the intersection of the convex hull of $\iota(M)$ with $\mathcal{B}$. Now X is a convex subset of a hyperbolic metric space, so is itself a hyperbolic metric space. Essentially by definition, $\partial X = \iota(M) \cong M$. It remains to verify that X is proper.

The convex hull (in the linear sense) of a compact subset of a Banach space is compact, so that $X \cup \iota(M)$ is compact. As $\iota(M)$ is closed in this set, this implies that X is locally compact. Finally, $\mathcal{B}$ is complete with respect to the Cayley-Klein metric, and X is convex (therefore closed) in $\mathcal{B}$, so X is also complete. The Hopf-Rinow theorem now tells us that X is a proper metric space. $\square$

There are more general constructions, often called *hyperbolic cones* which allows one to realise any complete and bounded metric space as the boundary of a hyperbolic metric space: see Chapter 6 of Buyalo–Schroeder.

PROPOSITION 4.14. *Let X and Y be proper hyperbolic metric spaces. If $f\colon X \to Y$ is a quasi-isometric embedding, then there is an induced map $\partial f\colon \partial X \to \partial Y$ which is a topological embedding. If, further, f is a quasi-isometry, then ∂f is a homeomorphism.*

PROOF. Let $\lambda \geq 1$ and $c \geq 0$ be quasi-isometry constants for f. We will define ∂f by pushing forward a geodesic representative of each point in ∂X. Let $p \in \partial X$ and let $\gamma\colon [0,\infty) \to X$ be a geodesic ray based at a point $x \in X$ that tends to p. By definition, the path $f \circ \gamma$ is a (λ, c)-quasigeodesic ray in Y. By the Morse Lemma, there is a constant $M \geq 0$ such that for each $t \in (0,\infty)$, the segment $f \circ \gamma|_{[0,t]}$ is a Hausdorff distance of at most M from a geodesic ξ_t with the same endpoints. Moreover, as f is a quasi-isometry, $\mathrm{d}(f(y), f(\gamma(t))) \to \infty$ as $t \to \infty$.

By properness and the Arzelà–Ascoli theorem, the sequence of paths $(\xi_n)_{n\in\mathbb{N}}$ has a subsequence converging to a geodesic ray $\xi\colon [0,\infty) \to Y$ based at $y = f(x)$, with endpoint $q \in \partial Y$. We define $\partial f(p) = q$. Since any two geodesics define the same point in ∂X if and only if they lie within a finite Hausdorff distance of one another, the same is true of their images under f. It follows that the map ∂f is well-defined and injective.

Let us show that ∂f is continuous. Let $q \in \mathrm{im}(\partial f)$ and $s > 0$, and let δ be a hyperbolicity constant for X and Y. We need to exhibit $r > 0$ such that $\partial f(U(p,r)) \subseteq U(q,s)$, where $\partial f(p) = q$. Take

$$r = (s + M + \frac{1}{2}\lambda\delta + \frac{1}{2}c)\lambda + c + M + 1$$

and let $p' \in U(p,r)$. Let γ and γ' be geodesic rays based at a point $x \in X$ tending to p and p' respectively. By hyperbolicity and the definition of $U(p,r)$, we have $\mathrm{d}_X(\gamma(t), \gamma'(t)) \leq \delta$ for $t \leq r$. This gives us

$$\mathrm{d}_Y(f\gamma(t), f\gamma'(t)) \leq \lambda\delta + c \quad \text{for } t \leq r \tag{4.1}$$

Now $f \circ \gamma$ and $f \circ \gamma'$ are (λ, c)-quasigeodesic rays based at $y = f(x) \in Y$. By the Morse Lemma and arguments similar to above, therefore, there is $M \geq 0$ such that they are M-close to geodesics ξ and ξ' based at y and tending to $q = \partial f(p)$ and $q' = \partial f(p')$

respectively. It then follows from (4.1) then

$$\mathrm{d}_Y(\xi(t), \xi'(t)) \leq 2M + \lambda\delta + c \quad \text{for } t \leq \frac{1}{\lambda} r - c - M.$$

We can use this to bound the inner product from below:

$$\begin{aligned} 2\langle \xi(t), \xi'(t) \rangle_y &= \mathrm{d}(\xi(t), y) + \mathrm{d}(\xi'(t), y) - \mathrm{d}(\xi(t), \xi'(t)) \\ &\geq 2t - 2M - \lambda\delta - c \end{aligned}$$

for $t \leq \frac{1}{\lambda} r - c - M$. In particular, by choice of r, taking $t = \frac{1}{\lambda} r - c - M > 0$ shows us that $\langle \xi(t), \xi'(t) \rangle_y \geq s$. Observing that the inner product is monotone increasing in both factors along geodesics shows that $q' \in U(q, s)$ as required.

When f is a quasi-isometry, it has a quasi-inverse $g\colon Y \to X$ which satisfies $\mathrm{d}_\infty(g \circ f, \mathrm{id}_X) < \infty$. It follows from the functoriality properties of the next exercise that

$$\partial g \circ \partial f = \partial (g \circ f) = \partial \, \mathrm{id}_X = \mathrm{id}_{\partial X},$$

so that necessarily ∂f is a surjective map. □

The homeomorphism type of the boundary can serve as a useful quasi-isometry invariant. For example, one can distinguish the real hyperbolic spaces $\mathbb{H}^n$ and $\mathbb{H}^m$ from one another up to quasi-isometry when $n \neq m$, as their boundaries are spheres of different dimensions.

Note that the boundary is not a complete invariant: real hyperbolic $2n$-space $\mathbb{H}^{2n}$ and complex hyperbolic n-space $\mathbb{H}^n_{\mathbb{C}}$ (the unique simply connected Hermitian manifold with constant negative holomorphic sectional curvature) are not quasi-isometric, though both have Gromov boundary S^{2n-1}. The *quasi-conformal structure* of the boundary of a hyperbolic space, a finer structure than its topological type, is actually enough to recover the hyperbolic space up to quasi-isometry, though we will not prove this here.

Exercise 4.15. Let X, Y, and Z be proper hyperbolic metric spaces. Show that the operator ∂ satisfies the following properties:

(1) If $f, g\colon X \to Y$ are quasi-isometric embeddings with $\mathrm{d}_\infty(f, g) < \infty$, then $\partial f = \partial g$;
(2) If $f\colon X \to Y$ and $g\colon Y \to Z$ are quasi-isometric embeddings, then $\partial(g \circ f) = \partial g \circ \partial f$;

In other words, ∂ is a functor from the category of proper hyperbolic metric spaces (with quasi-isometric embeddings as morphisms) to the category of compact metrisable spaces (with quasi-symmetric embeddings as morphisms).

5. Isometries of hyperbolic metric spaces

In this section we will classify isometries of hyperbolic metric spaces. We will see that a similar trichotomy holds as for those of classical hyperbolic spaces. Below is a largely elementary proof, stolen from the book of Coornaert–Delzant–Papadopoulos. There will be an arguably simpler proof of this fact later on, using much more advanced machinery.

Definition 5.1. Let X be a hyperbolic metric space, and $g \in \mathrm{Isom}(X)$ an isometry. We say that g is *elliptic* if the set $\{g^n x\}_{n \in \mathbb{N}}$ is bounded for any $x \in X$, that g is *parabolic* if g has exactly one fixed point in ∂X, and that g is *loxodromic* if it has exactly two fixed points in ∂X.

EXERCISE 5.2. Let X be a proper hyperbolic metric space. Show that $g \in \mathrm{Isom}(X)$ is loxodromic if and only if the orbit map $n \mapsto g^n x$ is a quasi-isometry $\mathbb{Z} \to X$ for any $x \in X$.

(Hint: Consider the geodesic lines between the two fixed points of g. The set of these lines is preserved by g: show that g acts as a sort of translation along these lines. Picking a point on these lines realises the orbit map as a quasi-isometry (where the constants are related to the translation length).

THEOREM 5.3. *Every isometry of a hyperbolic metric space is either elliptic, parabolic, or loxodromic.*

To show this, we will require the following two lemmas, which give us criteria for certain isometries of hyperbolic metric spaces to be loxodromic.

LEMMA 5.4. *Let X be a δ-hyperbolic metric space, $g \in \mathrm{Isom}(X)$ an isometry. If there is $x \in X$ with*

$$\mathrm{d}(gx, x) > 2\langle g^2x, x\rangle_{gx} + 2\delta,$$

then g is a loxodromic isometry.

PROOF. For each $n \in \mathbb{N}$, we write $d_n = \mathrm{d}(g^n x, x)$. We may thus rewrite the lemma hypothesis as

$$d_2 \geq d_1 + 2\delta + \varepsilon \tag{5.1}$$

where $\varepsilon > 0$. We will show by induction that $d_n \geq d_{n-1} + \varepsilon$ for $n \in \mathbb{N}$. For the base case that $n = 1$ observe that by the triangle inequality, we have $d_2 \leq 2d_1$. Combining this with (5.1) implies $d_1 \geq \varepsilon$.

Of course, the case $n = 2$ follows from (5.1). Let $n \geq 3$ and suppose that $d_n \geq d_{n-1} + \varepsilon$. We apply the four-point condition to the points x, gx, g^2x, and $g^{n+1}x$, giving that $d_2 + d_n \leq \max\{d_1 + d_{n+1}, d_1 + d_{n-1}\} + 2\delta$. Rearranging, we have

$$d_2 + d_n - d_1 - 2\delta \leq \max\{d_{n+1}, d_{n-1}\}.$$

Rearranging the equation (5.1) reduces the above to

$$d_n + \varepsilon \leq \max\{d_{n+1}, d_{n-1}\}.$$

By the induction hypothesis, $d_{n-1} + \varepsilon \leq d_n$, so we are must in fact have $d + n + \varepsilon \leq d_{n+1}$, as required. It follows immediately that $d_n \geq n\varepsilon$. Now

$$\varepsilon|n - m| \leq \mathrm{d}(g^n x, g^m x) \leq d_1|n - m|$$

so that the map $n \to g^n x$ is a quasi-isometry, completing the lemma. □

LEMMA 5.5. *Let X be a δ-hyperbolic metric space, and $g, h \in \mathrm{Isom}(X)$ non-loxodromic isometries. If there is $x \in X$ such that*

$$\mathrm{d}(gx, x) > 2\langle gx, hx\rangle_x + 6\delta \quad \text{and} \quad \mathrm{d}(hx, x) > 2\langle gx, hx\rangle_x + 6\delta$$

then gh and hg are loxodromic isometries.

PROOF. Let $x \in X$ be as in the lemma statement. We will attempt to find a bound on the Gromov product $\langle x, ghghx\rangle_{ghx}$, so that we can apply Lemma 5.4 to the isometry gh. The proof is quite technical and involves a lot of inequalities, but the idea is simply that one can transfer the condition on the bounded inner products in the hypothesis along a polygon in X to get the required bound. The fact that X is hyperbolic is what allows one to do this without losing too much length, essentially.

Since g and h are not loxodromic, Lemma 5.4 tells us that

$$\mathrm{d}(g^2x, x) \leq \mathrm{d}(gx, x) + 2\delta \quad \text{and} \quad \mathrm{d}(h^2x, x) \leq \mathrm{d}(hx, x) + 2\delta. \tag{5.2}$$

Moreover, rearranging the lemma hypotheses, we obtain

$$\mathrm{d}(gx, hx) \geq \mathrm{d}(gx, x) + 6\delta + \varepsilon \quad \text{and} \quad \mathrm{d}(gx, hx) \geq \mathrm{d}(hx, x) + 6\delta + \varepsilon \tag{5.3}$$

for some $\varepsilon > 0$. Now applying the four-point condition to x, gx, g^2x, and ghx and simplifying, we obtain

$$\mathrm{d}(gx, x) + \mathrm{d}(gx, hx) \leq \max\{\mathrm{d}(g^2x, x) + \mathrm{d}(hx, x), \mathrm{d}(gx, x) + \mathrm{d}(ghx, x)\} + 2\delta.$$

But (5.2) and (5.3) imply that $\mathrm{d}(gx, x) + \mathrm{d}(gx, hx) \geq \mathrm{d}(g^2x, x) + \mathrm{d}(hx, x) + 4\delta + \varepsilon$. As $\varepsilon > 0$, we must have that $\mathrm{d}(gx, x) + \mathrm{d}(gx, hx) \leq \mathrm{d}(gx, x) + \mathrm{d}(ghx, x) + 2\delta$. A symmetrical argument applies for hg, considering the four points x, hx, h^2x, and hgx. Simplifying and applying (5.3), we get the equations

$$\begin{aligned} \mathrm{d}(gx, x) + 4\delta + \varepsilon &\leq \mathrm{d}(ghx, x) \quad \text{and} \quad \mathrm{d}(gx, x) + 4\delta + \varepsilon \leq \mathrm{d}(hgx, x), \\ \mathrm{d}(hx, x) + 4\delta + \varepsilon &\leq \mathrm{d}(ghx, x) \quad \text{and} \quad \mathrm{d}(hx, x) + 4\delta + \varepsilon \leq \mathrm{d}(hgx, x). \end{aligned} \tag{5.4}$$

Now applying the four-point condition to the points x, gx, ghx, and $ghgx$ gives similarly

$$\mathrm{d}(ghx, x) + \mathrm{d}(hgx, x) \leq \max\{2\,\mathrm{d}(gx, x), \mathrm{d}(x, ghgx) + \mathrm{d}(hx, x)\} + 2\delta.$$

Now the first line of (5.4) shows that $\mathrm{d}(ghx, x) + \mathrm{d}(hgx, x) > 2\,\mathrm{d}(gx, x)$, so the first term in the above maximum is redundant. Two applications of (5.4) then tell us that we have

$$\mathrm{d}(gx, x) + 6\delta + \varepsilon \leq \mathrm{d}(ghgx, x) \quad \text{and} \quad \mathrm{d}(hx, x) + 6\delta + \varepsilon \leq \mathrm{d}(ghgx, x). \tag{5.5}$$

Finally, we apply the four-point condition to the points $x, ghx, ghgx$, and $ghghx$. This gives us

$$\mathrm{d}(ghx, x) + \mathrm{d}(ghgx, x) \leq \max\{\mathrm{d}(ghx, x) + \mathrm{d}(hx, x), \mathrm{d}(ghghx, x) + \mathrm{d}(gx, x)\} + 2\delta.$$

Similarly to before, (5.5) rules out the first term in the maximum. Applying (5.5) to the remaining inequality gives us

$$\mathrm{d}(ghx, x) + 6\delta + \varepsilon \leq \mathrm{d}(ghghx, x),$$

which one straightforwardly rearranges to see that $\mathrm{d}(ghx, x) > \langle ghghx, x \rangle_{ghx}$. We now apply Lemma 5.4 to conclude that gh is loxodromic. A symmetrical argument concludes the same about hg. ☐

Proof of Theorem 5.3. Let $g \in \mathrm{Isom}(X)$ be an isometry that is not elliptic, parabolic, or loxodromic, and let $x \in X$. As g is not elliptic, the orbit $\{g^n x\}_{n \in \mathbb{N}}$ is unbounded. Since $X \cup \partial X$ is compact by Proposition 4.12 and g is not parabolic, there are subsequences $(g^{n_i}x)$ and $(g^{m_i}x)$ converging to distinct points $a, b \in \partial X$.

By the definition of the topology on ∂X, there is some $r \geq 0$ and $N \in \mathbb{N}$ such that

$$\langle g^{n_i}x, g^{m_i}x \rangle_x \leq r$$

for all $i \geq N$. Now since g has unbounded orbits, there is $N' \in \mathbb{N}$ such that $\mathrm{d}(g^n x, x) \geq 2r + 6\delta$ for $n \geq N'$. Choosing $i \geq \max\{N', n_N, m_N\}$ allows us to apply Lemma 5.5 to see that $g^{n_i + m_i}$ is loxodromic. This contradicts the fact that g is not loxodromic. ☐

CHAPTER 3

Negative curvature in groups

In this chapter we will apply the machinery we have developed for metric negative curvature to group theory, by studying the groups that act on hyperbolic metric spaces. The most well-behaved such groups are, naturally, those that act geometrically on proper hyperbolic metric spaces. Remarkably, this entirely geometric condition has incredibly strong consequences for the algebraic structure of these groups.

1. Hyperbolic groups

The class of hyperbolic groups includes many groups of classical interest to group theorists, topologists, and geometers. As well as providing a convenient and clarifying framework for understanding groups with 'negative curvature', the study of hyperbolic groups has paved the way for some deep insights into and novel results on these groups.

Definition 1.1 (Hyperbolic group). A group is called *hyperbolic* if it admits a geometric action on a proper hyperbolic metric space.

Equivalently, we may say a finitely generated group is hyperbolic if it has a finite generating set with respect to which the Cayley graph is a a hyperbolic metric space, invoking the Milnor–Schwarz lemma. By Lemma 3.6, the quasi-isometry type of a Cayley graph is preserved by a change of finite generating set, and hyperbolicity is preserved by quasi-isometries by Corollary 2.15. Hence the hyperbolicity of any such Cayley graph is independent of which finite generating set is chosen for a hyperbolic group.

Example 1.2.

- The Cayley graph of any finitely generated free group with respect to a free generating set is a simplicial tree, and hence 0-hyperbolic. Therefore finitely generated free groups are hyperbolic.
- If M is a closed hyperbolic n-manifold, its fundamental group $\pi_1 M$ acts geometrically on its isometric universal cover $\mathbb{H}^n$. We saw earlier that $\mathbb{H}^n$ is a hyperbolic metric space, so $\pi_1 M$ is a hyperbolic group.
- Every finite group is hyperbolic as it is quasi-isometric to a point, and every virtually cyclic group is hyperbolic as it is quasi-isometric to a line. We call these *elementary* hyperbolic groups – they are the only virtually abelian ones – and all others *non-elementary*. We will later see that all non-elementary hyperbolic groups contain non-abelian free subgroups, so are very far from being virtually abelian.
- So-called 'random groups' are hyperbolic. More precisely, one can formulate models of randomness that allow one to choose a finite presentation 'uniformly

randomly' in some sense. In most of these models, the 'generic' group is almost always a hyperbolic group.
- A group given by a presentation with relators that do not overlap too much is hyperbolic. Such 'small cancellation' groups are a rich source of examples in geometric group theory, and can exhibit somewhat peculiar properties. We will discuss this class of groups a little later in the course.

Example 1.3. The Cayley graph of $\mathbb{Z}^n$ with respect to the standard generators is not a hyperbolic metric space for any $n \geq 2$, and so $\mathbb{Z}^n$ is not a hyperbolic group.

We will see there is a sort of strong converse to the above example, in that hyperbolic groups cannot contain higher rank abelian groups. This, among other things, will be a consequence of the following important fact.

Theorem 1.4. *Every infinite order element of a hyperbolic group is loxodromic.*

Proof. Let S be a finite generating set for the group G such that $\Gamma(G,S)$ is δ-hyperbolic, and let g be an element of infinite order. Let N be the number of group elements h with $|h|_S \leq 2\delta$, of which there are finitely many. We will show that for any $R \in \mathbb{N}$, we have $\left|g^{RN}\right|_S \geq R$. It follows immediately that g^N is a loxodromic, as it implies

$$|n-m| \leq \mathrm{d}_S(g^{nN}, g^{mN}) \leq |g|_S \cdot |n-m|.$$

That g is a root of a loxodromic then implies g is a loxodromic, so it remains only to prove the claim.

Let $R \in \mathbb{N}$ and take $k \in \mathbb{N}$ large enough so that $\left|g^k\right|_S \geq 4R+2\delta+1$. Now if $|g^n|_S \leq R$, then the geodesic $[1, g^k]$ and its g^n-translate $[g^n, g^{n+k}]$ are at length at least $4R+2\delta$ and have endpoints at most R apart. We leave it as a straightforward exercise in hyperbolic geometry that the midpoint of the latter path is a distance of at most 2δ from a point on the former path that is at most $\frac{1}{2}R$ away from the midpoint of the former path. There are at most RN such points, by the definition of N. As g does not fix any points in $\Gamma(G,S)$, the midpoints of $[g^n, g^{n+k}]$ must all be distinct. By the pigeonhole principle, then, there is some $n(R) \leq RN+1$ with $\left|g^{n(R)}\right|_S > R$. It follows also that $R \leq n(R)|g|_S$.

Suppose now that $\left|g^{RN}\right| \leq R - \varepsilon$ for some $\varepsilon > 0$. Let $T = \max\{\left|g^i\right|_S \,|\, 0 \leq i < RN\}$ and $N' = \lceil \frac{1}{\varepsilon} RNT \rceil$. Then for any $n \geq N'$, we have

$$|g^n|_S \leq \left|g^{RN}\right|_S^p + |g^q|_S \leq pR - p\varepsilon + T \leq pR,$$

where $p, q \in \mathbb{Z}$ are such that $n = pRN + q$ and $0 \leq q < RN$. Let $Q = N'|g|_S$, so that $n(Q) \geq N'$. It follows that $\left|g^{n(Q)}\right|_S \leq \frac{n(Q)}{RN} R \leq Q$, while the construction of $n(Q)$ gives that $\left|g^{n(Q)}\right|_S > Q$. This is a contradiction, so we must have $\left|g^{RN}\right| \geq R$ □

The geometric condition of hyperbolicity has some strong implications for the algebraic structure of the group. The beginning of this study sees that centralisers of infinite order elements are always virtually cyclic.

Theorem 1.5. *Let G be a hyperbolic group. If $g \in G$ is an element of infinite order, then $[C_G(g) : \langle g \rangle] < \infty$.*

PROOF. Take S be a finite generating set for G, so that $\Gamma(G, S)$ is δ-hyperbolic. Let $\lambda \geq 1$ and $c \geq 0$ be constants for which $n \to g^n$ is a (λ, c)-quasi-isometry $\mathbb{Z} \to \Gamma(G, S)$. We consider this map a quasi-geodesic by precomposing it with a quasi-isometry $\mathbb{R} \to \mathbb{Z}$. Let $M = M(\lambda, c, \delta)$ be the constant obtained by the Morse Lemma.

Let $h \in C_G(g)$ be an arbitrary element of the centraliser of g and write $D = |h|_S$. Since g has loxodromic, there is $N \in \mathbb{N}$ such that $\mathrm{d}_S(1, g^n) > 2\delta + 2M + D$ for all $n \geq N$. Let $n \geq N$ and choose geodesics $p_1 = [1, g^{2n}], p_2 = [h, hg^{2n}], q_1 = [1, h]$, and $q_2 = [g^{2n}, hg^{2n}]$. We may take p_2 to be a h-translate of p_1. These four geodesics form a geodesic rectangle in $\Gamma(G, S)$, which is 2δ-slim as $\Gamma(G, S)$ is δ-hyperbolic.

By Theorem 1.4, the points $\{1, g, \ldots, g^{2n}\}$ are the image of a (λ, c)-quasi-geodesic. Therefore by the Morse Lemma, they lie in an M-Hausdorff neighbourhood of p_1. As $p_2 = hp_1$, the same is true for $\{h, hg, \ldots, hg^{2n}\}$ and p_2. Let y_1 be a point on p_1 with $\mathrm{d}_S(y_1, g^n) \leq M$. By the choice of n, we have $\mathrm{d}_S(y_1, q_i) > 2\delta$ for $i = 1, 2$. Therefore by the slimness of the rectangle, there is a point y_2 be a point on p_2 with $\mathrm{d}_S(y_1, y_2) \leq 2\delta$. Now there some index $j = 0, \ldots, 2n$ such that $\mathrm{d}_S(y_2, hg^j) \leq M$.

Combining all of this, we have $\mathrm{d}_S(hg^j, g^n) \leq 2M + 2\delta$. Using that h commutes with g, this implies $\mathrm{d}_S(h, g^{n-j}) < 2M + 2\delta$. In other words, $h \in a\langle g\rangle$, where $a \in G$ is such that $|a|_S \leq 2M + 2\delta$. As S is a finite set, there are only finitely many such elements. Thus $\langle g\rangle$ has finite index in $C_G(g)$ as required. □

An immediate consequence of this is that hyperbolic groups contain no subgroups isomorphic to the Baumslag-Solitar group $\mathrm{BS}(n, n) = \langle a, b \mid ba^n b^{-1} = a^n\rangle$, for the whole group centralises the infinite order element a^n. In particular, hyperbolic groups cannot contain any higher rank abelian subgroups, as $\mathbb{Z}^2 \cong \mathrm{BS}(1, 1)$.

Another algebraic consequence of hyperbolicity is that one has strong control over the torsion elements of the group. We examine a simple case to get an intuition for why one should be able to draw such conclusions.

EXAMPLE 1.6. Let G be a group acting geometrically on a simplicial tree T (that is, a 0-hyperbolic graph), and let $H \leqslant G$ be a finite subgroup. As H is finite, the orbit Hx of any point $x \in T$ is a finite set. Thus H fixes the barycentre of Hx; it is a subgroup of a point stabiliser. Since the action is cocompact, there are finitely many conjugacy classes of point stabilisers. Moreover, since the action is proper, each point stabiliser is finite. It follows that there are only finitely many conjugacy classes of finite subgroups in G.

The general idea of the above example generalises to the hyperbolic of groups acting geometrically on hyperbolic spaces, with some complications. In trees, it is easy to define a centre for a finite set of points, while this is not so obvious in hyperbolic spaces in general.

THEOREM 1.7. *Hyperbolic groups contain finitely many conjugacy classes of finite subgroups.*

PROOF. Let G be a hyperbolic group with a geometric action on a δ-hyperbolic metric space X. Let $H \leqslant G$ be a finite subgroup. We will show that H preserves a *quasi-centre* of its orbits. For a bounded subset $Y \subseteq X$, denote

$$R_Y = \inf\{r > 0 \mid Y \subseteq B_r(x) \text{ for some } x \in X\},$$

and define the set
$$C(Y) = \{x \in X \mid Y \subseteq B_{R_Y+1}(x)\}.$$
This set is non-empty by definition of R_Y. We claim that $\operatorname{diam}(C(Y)) \leq 4\delta + 2$.

Let $x, x' \in C(Y)$, and let m be the midpoint of a geodesic $[x, x']$. Let $y \in Y$ be an arbitrary point in Y. By hyperbolicity, there is a point t on $[x, y]$ or $[x', y]$ with $\mathrm{d}(m, t) \leq \delta$. Suppose without loss of generality that it is the former. Now
$$\begin{aligned} \mathrm{d}(y, m) &\leq \mathrm{d}(y, t) + \mathrm{d}(t, m) \\ &\leq \mathrm{d}(y, x) - \mathrm{d}(x, t) + \delta \leq R_Y + 1 + 2\delta - \mathrm{d}(x, m). \end{aligned}$$
On the other hand, there must be some $y \in Y$ with $\mathrm{d}(y, m) \geq R_Y$. Rearranging the above equation for this y gives $\mathrm{d}(x, m) \leq 2\delta + 1$. As m is the midpoint of $[x, x']$, the claim follows.

Fix a point $x \in X$, and let $B \subseteq X$ be a compact subset such that $G \cdot B = X$, which exists as the action is cocompact. Write $K = N_{4\delta}(B)$ and note that K is also compact as X is proper. As the action is properly discontinuous, the set $T = \{g \in G \mid gK \cap K \neq \emptyset\}$ is finite. Thus T contains finitely many distinct subgroups.

The orbit Hx is a bounded subset of X. As the orbit Hx is setwise preserved by H, the quasi-centre $C(Hx)$ is also setwise preserved by H. Moreover, there is some $g \in G$ such that $gC(Hx) \cap B \neq \emptyset$, since $G \cdot B = X$. Thus gHg^{-1} setwise fixes the translate $gC(Hx)$. By the claim $C(Hx)$ is a set of diameter at most $4\delta + 2$ containing the identity, which implies that $gC(Hx) \subseteq K$. Therefore $gHg^{-1} \subseteq T$, completing the theorem. $\square$

Hyperbolicity also allows one to rule out certain pathologies. A group in which all elements have finite order is often called a *torsion group*, or a *periodic group*. One pathology one might consider is that of being infinite while also having no elements of infinite order, that is, being an infinite torsion group. Of course, there are many silly examples of infinite torsion groups, such as an infinite direct product of finite groups, the quotient group $\mathbb{Q}/\mathbb{Z}$, or the Prüfer group $\mathbb{Z}(p^\infty)$, but sensible groups generally tend not to contain these. We will defer the proof until later on.

THEOREM 1.8. *A hyperbolic group contains no infinite torsion subgroups.*

REMARK 1.9. The existence of finitely generated infinite torsion groups was a for a long time a major open problem in group theory known as the *general Burnside problem.* After standing for over 60 years, a negative solution was given by Golod and Shafarevich in 1964. The groups they constructed arose in connection with the class field tower problem in number theory: they were interested in the infinitude of certain pro-p groups arising as Galois groups of certain extensions. They established a bound that relates the minimal number of relators and minimal number of generators for a finite p-group.

Another important property of hyperbolic groups is that they are, in a precise sense, very large and have many quotients. This is captured more exactly by the following theorem due to Ol'shanskii, which we is beyond the scope of this course.

THEOREM 1.10. *Let G be a non-elementary hyperbolic group. For any countable group C, there is a normal subgroup $N \lhd G$ with C is isomorphic to a subgroup of G/N.*

The property above is known as *SQ-universality*, and it satisfied by many of the generalisations of hyperbolic groups as well.

2. Convergence groups

We have already seen that boundaries provide a useful invariant for the coarse geometric features of a hyperbolic space. When it comes to groups, it turns out that one can further study the dynamical properties of the action of the group on its boundary to recover algebraic information about the group. In this section we will build a framework for understanding hyperbolic groups through their actions on boundary spaces.

2.1. Definitions and basic properties. If G is a group acting by isometries on a hyperbolic space X, every element of G induces a homeomorphism of its boundary ∂X. This gives us a representation $G \to \mathrm{Homeo}(\partial X)$. To describe what sort of action this is, we will take a step outwards in terms of the level of abstraction, and consider convergence groups: groups that act by homeomorphisms on arbitrary metrisable compacta. The motivation for the following definition really comes from the action of subgroups of $\mathrm{Isom}(\mathbb{H}^n)$ acting on the boundary sphere of $\mathbb{H}^n$. Indeed, the setting of subgroups of $\mathrm{Homeo}(S^n)$ arising from isometry groups of hyperbolic space is the origin of the notion of a convergence group.

Definition 2.1 (Convergence sequence). Let M be a compact metrisable space, and G a group acting on M by homeomorphisms. A sequence of elements $(g_n)_{n\in\mathbb{N}}$ of G is called a *convergence sequence* if there are points $a, b \in M$ such that g_n converges locally uniformly on $M - \{b\}$ to the constant function on a.

We call the points a and b the *attracting* and *repelling* points for the sequence (g_n). Note that a and b need not be distinct.

Exercise 2.2. Show that if (g_i) is a convergence sequence with attracting point a and repelling point b, then (g_i^{-1}) is a convergence sequence with attracting point b and repelling point a.

Definition 2.3 (Convergence group). Let M be a compact metrisable space, and G a group acting on M by homeomorphisms. We say that G is a *convergence group on M* if every infinite sequence of distinct elements has a convergence subsequence.

The *limit set* of G is the set ΛG of accumulation points of G-orbits of points in M. We say that the action is *minimal* if $\Lambda G = M$. Further the action is *elementary* if ΛG has at most two points, and is *non-elementary* otherwise.

Example 2.4.

- A finite group is a convergence group on the empty set, and every group is a convergence group on a point or two points with the trivial action.
- Let G be a group acting properly discontinuously on a hyperbolic space X. We will see later that G is a convergence group on ∂X.

Exercise 2.5. Suppose G is a convergence group on compact metrisable space M. Show that if G is non-elementary, then ΛG is uncountable.

Remark 2.6. It follows from the definition of a convergence group that the map $G \to \mathrm{Homeo}(M)$ has finite kernel, if G is a non-elementary convergence group on M.

Remark 2.7. The above definition is sometimes referred to as a *discrete convergence group* in the literature. A 'general' convergence sequence (g_i) is a sequence that is either a convergence sequence in the above sense, or otherwise converges uniformly to a homeomorphism $g \in \mathrm{Homeo}(M)$; a 'general' convergence group is a group that acts on a compact metrisable space and every infinite sequence has a general convergence subsequence. This notion is of interest as it allows one to study non-discrete group actions on compact spaces, but here we will only be interested in discrete groups.

We will engage in an analysis of the elements of convergence groups by studying fixed points and limit points on the boundary. Firstly, as with isometries of hyperbolic spaces, elements of convergence groups fall into a familiar trichotomy.

DEFINITION 2.8 (Elements of convergence groups). Let G be a convergence group on compact metrisable space M, and let $g \in G$ be an element. Then we say g is:

(i) *elliptic* if it has finite order;
(ii) *parabolic* if it is of infinite order and has exactly one fixed point in M; or
(iii) *loxodromic* if it is of infinite order and has exactly two fixed points in M.

LEMMA 2.9. *Every element of a convergence group is either elliptic, parabolic, or loxodromic.*

PROOF. Let G be a convergence group on M and $g \in G$ an infinite order element. Then there is some sequence (g^{n_i}) that is a convergence sequence with attracting point a and repelling point b. Now $g^{n_i}(gp) = gg^{n_i}p \to ga$ uniformly away from $g^{-1}b$, so that ga is also an attracting point for (g^{n_i}). Hence $ga = a$, so g has a fixed point in M. Moreover, since the sequence convergence to a constant function on a locally uniformly outside $M - \{b\}$, the only possible fixed points of g are a and b. Therefore any infinite order element is either parabolic or loxodromic. □

We have a dynamical criterion for being a loxodromic element. It allows us to reduce the rather exact property of having fixed points to a topological nesting property that is easier to verify.

LEMMA 2.10. *Let G be a convergence group on M and $g \in G$ an element. If there is a proper open subset $U \subseteq M$ such that $g\overline{U} \subseteq U$, then (g^i) is a convergence sequence with attracting point in U and repelling point in $M - \overline{U}$. Moreover, g is a loxodromic element.*

PROOF. Consider the sets $A = \bigcap g^n\overline{U}$ and $B = \bigcap g^{-n}(M - U)$. It is immediate from the definition that A and B are fixed by g. By definition, A and B are disjoint. We show that A and B both consist of singletons, which proves the lemma.

As G is a convergence group, there is a sequence (n_i) such that (g^{n_i}) is a convergence sequence with attracting point a and repelling point b. Of course, we must have $a \in A$ and $b \in B$. Suppose that $c \in B$ with $c \neq b$. Then $g^{n_i}c \in U$ for sufficiently large i. However, B is disjoint from U, so this is a contradiction. Hence $B = \{b\}$ and by a symmetrical argument with (g^{-n_i}) shows $A = \{a\}$. □

EXERCISE 2.11. Use the previous lemma to show that if (g_i) is a convergence sequence with distinct attracting and repelling points, the elements g_i are eventually loxodromic.

LEMMA 2.12. *Let G be a convergence group on M, and let $g \in G$ be an infinite order element. Then (g^i) is a convergence sequence whose attracting and repelling points coincide with the fixed points of g.*

Proof. Let (g^{n_i}) be a convergence subsequence of (g^i) with attracting point a and repelling point b. As in the proof of Lemma 2.9, a and b are necessarily fixed points of g. Suppose g is parabolic and that (g^i) is not a convergence sequence. Then there is a sequence of points $p_i \in M$ such that $p_i \to p \neq a$, and $g^{n_i}p_i \to q \neq a$, after possibly passing to a subsequence of (g^{n_i}). But then (g^{n_i}) has attracting point a and repelling point $p \neq a$. This implies that p and a are distinct fixed points of g, which contradicts the fact that g is parabolic.

Now suppose g is loxodromic. Then pick some neighbourhood U of a with $b \notin \overline{U}$. The set $\overline{U}$ is a compact subset not meeting b. As (g^{n_i}) is a convergence sequence, there is some n_i such that $g^{n_i}\overline{U} \subseteq U$. Now by Lemma 2.10, the element $h = g^{n_i}$ is loxodromic. Moreover, (h^j) is a convergence sequence with attracting point a and repelling point b. It follows that (g^i) is a convergence sequence also, since h is a power of g. $\square$

In light of the above, it makes sense to give special name to the fixed points corresponding to an element of a convergence group.

Definition 2.13 (Poles). Let G be a convergence group on M and $g \in G$ an infinite order element. We write P_g (respectively, N_g) for the attracting (respectively, repelling) point of the convergence sequence (g^i), and we call it the *positive* (respectively, *negative*) *pole of* g.

Of course, for a parabolic element, the positive and negative poles are the same point.

In hyperbolic space, if two axes share a single point at infinity, then translations along those axes do not generate a discrete subgroup of isometries. This behaviour is reflected in the discrete nature of convergence groups (cf. Remark 2.7), in the form that two elements cannot share one pole without sharing the other.

Lemma 2.14. *Let G be a convergence group on M, and $g, h \in G$ infinite order elements. Then the fixed point sets of g and h in M are either disjoint or coincide.*

Proof. If both g and h are parabolic, the statement is trivial. Suppose that both g and h are loxodromic, and suppose that $P_g = P_h$ while $N_g \neq N_h$. Let U be a neighbourhood of P_g such that $N_g, N_h \notin \overline{U}$. By Lemma 2.12, (g^i) and (h^i) are convergence sequences with attracting point P_g and repelling points N_g and N_h respectively. Then there are $i, j \in \mathbb{N}$ such that $g^i\overline{U}$ and $h^j\overline{U}$ are contained in U. For conveience, we relabel so that $g = g^i$ and $h = h^j$.

Define $F = \overline{U} - gU \neq \emptyset$, and note that the sets g^iF cover $U - \{P_g\}$. Let $p \in F$ be a point. Then for each i, there is n_i such that $h^ip \in g^{n_i}F$. Necessarily, $n_i \to \infty$ as $i \to \infty$. Let $k_i = g^{-n_i}h^i$, so that k_i fixes P_g, and $k_iN_h = g^{-n_i}N_h \to N_g$. This implies that there are infinitely many distinct k_i, since $N_g \neq N_h$. Let (k_{m_i}) be a convergence subsequence. By the above, the attracting and repelling points of (k_{m_i}) must be among P_g, N_g, and N_h. However, $k_i(p) \in F$ for all i and $P_g, N_g, N_h \notin \overline{F}$, a contradiction.

Finally, if g is loxodromic and h is parabolic, then g and hgh^{-1} are loxodromic elements with one shared fixed point, which we have just shown to be impossible. $\square$

Exercise 2.15. Show that if g is a loxodromic element of non-elementary convergence group G, then the stabiliser of $\mathrm{fix}(g)$ contains $\langle g \rangle$ has a finite index subgroup.

EXERCISE 2.16. Suppose that G is a non-elementary convergence group, $H \leqslant G$ a subgroup. Show that if ΛH consists of two points, then H is contains a loxodromic element g such that $\langle g \rangle$ has finite index in H.

2.2. Ping-pong and the Tits' alternative. The following gives us a way to construct a loxodromic with certain prescribed fixed points out of two others, a fact which will have some powerful algebraic consequences.

LEMMA 2.17. *Let G be a convergence group on M, and suppose that $g, h \in G$ are loxodromic elements with disjoint fixed point sets. Let U and V be neighbourhoods of P_g and P_h respectively. Then there is $n \geq 1$ such that $k = g^n h^{-n}$ is loxodromic with $P_k \in U$ and $N_k \in V$.*

PROOF. Let U_+, U_-, V_+, and V_- be neighbourhoods of P_g, N_g, P_h, and N_h respectively, whose closures are disjoint. We may suppose that $U_+ \subseteq U$ and $V_+ \subseteq V$. For i sufficiently large, we have the inclusions

$$\begin{aligned} g^i(M - U_-) &\subset U_+ \quad \text{and} \quad g^{-i}(M - U_+) \subset U_-, \\ h^j(M - V_-) &\subset V_+ \quad \text{and} \quad h^{-j}(M - V_+) \subset V_-. \end{aligned} \tag{2.1}$$

Let $k = g^i h^{-i}$ and observe that (2.1) implies that

$$k\overline{U_+} \subseteq g^i h^{-i}(M - V_+) \subseteq g^i V_- \subseteq g^i(M - U_-) \subseteq U_+.$$

Similarly, $k^{-1}\overline{V_+} \subseteq V_+$. Applying Lemma 2.10, we see that k is a loxodromic element whose fixed points lie in $U_+ \subseteq U$ and $V_+ \subseteq V$, as required. $\square$

The following criterion allows us to build free groups of a given rank in a convergence group, given certain loxodromic elements. The general idea is to find some elements that bat some disjoint subsets of the space back and forth (hence the 'ping-pong'), and these elements necessarily generate a free group as a result. Ping-pong type arguments apply in many settings, and were initially utilised by Jacques Tits, who used them to prove that every subgroup of a finitely generated linear group either has a finite index solvable subgroup, or contains a non-abelian free group. We will see shortly that a sort of analogue of this theorem holds for convergence groups.

PROPOSITION 2.18 (Ping-pong lemma). *Let G be a convergence group on compact metrisable space M. If $g_1, \ldots, g_n \in G$ are loxodromic elements with disjoint fixed point sets in M, then there are $m_1, \ldots, m_n \in \mathbb{Z}$ such that $g_1^{m_1}, \ldots g_n^{m_n}$ freely generate a free group of rank n.*

PROOF. For convenience we treat the case that $n = 2$; the general argument is virtually identical. Let g and h be loxodromic elements of G with disjoint fixed point sets. Pick neighbourhoods U_+, U_-, V_+, and V_- of $\{P_g, N_g\}$ and $\{P_h, N_h\}$ respectively, whose closures are disjoint. Since (g^n) and (h^n) are convergence sequences, there are i and j such that (2.1) holds. For convenience, we may replace g with g^i and h with h^j.

Let $F = F(g, h)$ be the free group generated by g and h and consider the natural homomorphism $\varphi \colon F \to \langle g, h \rangle \leqslant G$. We must show that φ is injective.

For every word w in $A = \{g, h, g^{-1}, h^{-1}\}$, we inductively define the set $X(w) \subseteq M$ as follows. Call $U(g) = U_+, U(g^{-1}) = U_-, U(h) = V_+, U(h^{-1}) = V_-$. We define

$$X(a) = \varphi(a) \cdot \bigcup_{b \neq a^{-1}} U(b)$$

for any letter $a \in A$. Finally, we inductively define $X(aw) = \varphi(a) \cdot X(w)$ for any letter $a \in A$ and word w in A. Now let w be a reduced word in A with $\varphi(w) = 1$. If w is non-empty, then $X(w) \subseteq U(a)$ by (2.1), where a is the first letter of w. But this means that $\varphi(w)$ acts non-trivially on the sets $U(b)$, where $b \neq a^{-1}$, contradicting that $\varphi(w) = 1$. Hence w is the empty word, and φ is injective as required. $\square$

We show that this is in fact a generic situation in a non-elementary convergence group, in the sense that loxodromic fixed point pairs are dense in the limit set. As a preliminary, we first prove a one-sided version of this.

Lemma 2.19. *Let G be a convergence group on M such that ΛG has at least two points. Then for every open subset $U \subseteq M$ with non-empty intersection with ΛG, there is a loxodromic element of G with a fixed point in U.*

Proof. Let $U \subseteq M$ be an open subset, and pick limit points $a \in U$ and $a' \neq a$. Since $a, a' \in \Lambda G$, there are convergence sequences (g_i) and (h_i) whose attracting points are a and a' respectively. Let b and b' be the repelling points of (g_i) and (h_i). Let V be a neighbourhood of a' whose closure is disjoint from that of U, and $b, b' \notin V$ (if $b, b' \neq a'$).

If $a \neq b$, then g_i is loxodromic with one fixed point in U for sufficiently large i. Similarly, if $a' \neq b'$, then h_i is loxodromic with a fixed point in V for sufficiently large i. Since (g_j) converges locally uniformly on $M - \{b\}$ to the constant function on a, for sufficiently large j, the element $g_j h_i g_j^{-1}$ is a loxodromic with fixed point in $g_j a' \in U$.

It thus remains to consider only the case that $a = b$ and $a' = b'$. Then $h_i \overline{U} \subseteq V$ for sufficiently large i, and likewise $g_j \overline{V} \subseteq U$ for sufficiently large j. Hence $g_j h_i \overline{U} \subseteq U$, whence $g_j h_i$ is loxodromic with a fixed point in U. $\square$

Theorem 2.20. *Let G be a convergence group on M. For any disjoint open subsets U and V that have non-empty intersection with ΛG, there is a loxodromic element $g \in G$ with $P_g \in U$ and $N_g \in V$.*

Proof. By Lemma 2.19, there are loxodromic elements g and h with $P_g \in U$ and $P_h \in V$. If $N_g = P_h$ or $N_h = P_g$, then the fixed points of g and h coincide, and so g is a loxodromic with endpoints in U and V. Suppose otherwise then, that the fixed points of g and h are distinct from one another. Then by Lemma 2.17, $k = g^n h^{-n}$ yields a loxodromic element with $P_k \in U$ and $N_k \in V$, for large enough n. $\square$

The density of loxodromic fixed point pairs gives us the aforementioned dichotomy for subgroups of a convergence group.

Corollary 2.21 (Tits' alternative). *Let G be a convergence group, $H \leqslant G$ a subgroup. If H is non-elementary, it contains a non-abelian free subgroup.*

Proof. Since H is nonelementary, $\Lambda H \subseteq \Lambda G$ is uncountable. In particular, ΛH contains at least four points. Picking disjoint neighbourhoods of these, we can apply

Theorem 2.20 to find a pair of loxodromic elements with distinct poles in ΛH. The theorem now follows from Proposition 2.18. $\square$

2.3. Uniform convergence groups. We now analyse the behaviour of limit points in the limit set of a convergence group in depth. A particularly important class of limit points are the *conical limit points*, which are points that are well-approximated by orbits of the group.

Definition 2.22 (Conical limit point). Let G be a convergence group on compact metrisable space M. A point $p \in M$ is called a *conical limit point* if there are distinct points $a, b \in M$ and a sequence (g_i) in G such that $g_i p \to b$ and $g_i q \to a$ for all $q \in M - \{p\}$.

Example 2.23. If a and b are fixed points of a loxodromic element g, then the sequences (g^n) and (g^{-n}) are witnesses to the fact that a and b are conical limit points.

Replacing (g_i) in the above with a convergence subsequence and taking inverses, it is immediate that a conical limit point of a convergence group G is in fact a limit point of G (that is, it is contained in ΛG). The origin of 'conical' in the name above is in reference to a characterisation of such points in classical hyperbolic geometry. Suppose that $p \in \partial \mathbb{H}^n$ is a point in the boundary of $\mathbb{H}^n$, and $G \leqslant \mathrm{Isom}(\mathbb{H}^n)$ is a discrete group of isometries. A neighbourhood of a line in $\mathbb{H}^n$ tending to p is a exactly a cone in the upper half space model, and p is a conical limit point of G if and only if there is an infinite G-orbit contained in such a cone.

Definition 2.24 (Uniform convergence group). A convergence group on compact metrisable space M is called *uniform* if every point of M is a conical limit point.

Of course, a uniform convergence group is necessarily minimal. The power of the above definition lies in the fact that the dynamics of conical limit point is very constrained. That is, one has strong control over the types of elements and subgroups that fix conical limit points. For instance, in contrast with loxodromic fixed points, parabolic fixed points can never be conical limit points.

Lemma 2.25. *Let G be a convergence group on M. If p is a parabolic fixed point, then it is not a conical limit point. In particular, a uniform convergence group contains no parabolic elements.*

Proof. Suppose that $p \in M$ is both a parabolic fixed point and a conical limit point. Thus there is a parabolic element $g \in G$ with fixed point p and also a sequence (h_n) and distinct points $a, b \in M$ with $h_n p \to b$ and $h_n q \to a$ for $q \neq p$. We may suppose without loss of generality that $h_n p \neq a$ for any n, by deleting finitely many terms in the sequence.

Fix some $n \in \mathbb{N}$ and consider the sequence $(h_i^{-1} h_n)_{i \in \mathbb{N}}$. By definition, we have that $h_i^{-1} h_n q \to p$ for $q \neq h_n^{-1} a$. Passing to a convergence subsequence, $(h_i^{-1} h_n)$ has attracting point p and repelling point $h_n^{-1} a \neq p$. Hence the elements $h_i^{-1} h_n$ are eventually loxodromic.

Consider the sequence of elements $k_n = h_n g h_n^{-1}$ for $n \in \mathbb{N}$. Since k_n is a conjugate of a parabolic element, it is also parabolic. We claim that k_m and k_n are distinct for any n and sufficiently large m (and so the sequence has infinitely many distinct terms). Indeed, suppose otherwise, that $k_m = k_n$. This implies that $h_m g h_m^{-1} = h_n g h_n^{-1}$ or, to rephrase, that g commutes with $h_m^{-1}h_n$. However, for sufficiently large m, we saw that $h_m^{-1}h_n$ is loxodromic, and loxodromic fixed points cannot be parabolic fixed points, a contradiction. Hence we may pass to a convergence subsequence of (k_n).

Now since g fixes p, we have $k_n(h_n p) = h_n g p = h_n p \to b$. On the other hand, for any $q \neq p$, we have $h_n g q \to a$ and so $k_n(h_n q) = h_n g q \to a$. It follows that a and b are attracting and repelling points for k_n and so k_n is loxodromic for large enough n. However, then k_n and g share a single fixed point, which is a contradiction. $\square$

The following emphasises that conical points, when they actually are fixed points, behave like loxodromic fixed points. Indeed, one can think of conical limit points as points in M that want to be loxodromic fixed points. We will see later that non-elementary convergence groups are countable, and as such there can only be countably many loxodromic fixed points, while their limit sets are uncountable.

Proposition 2.26. *Let G be a convergence group on M, and suppose that $p \in M$ is fixed by infinitely many elements (h_j) of G. If p is a conical limit point, then h_j is loxodromic for some $j \in \mathbb{N}$.*

Proof. Let $a \neq b$ be points and (g_i) a sequence as in the definition of conical limit point, so $g_i p \to b$ and $g_i q \to a$ for $q \neq p$. We may pass to a convergence subsequence of (g_i); necessarily p is its repelling point and a is its attracting point.

Suppose first that for some $j \in \mathbb{N}$, there are infinitely many distinct conjugates $k_i = g_i h_j g_i^{-1}$. Thus we may pass to a convergence subsequence of (k_i). Fix some $q \neq p$, and define $p_i = g_i p$ and $q_i = g_i q$. We have that $k_i p_i = p_i \to b, q_i \to a$, and $k_i q_i = g_i h_j q \to a$, as $h_j q \neq p$. It follows that a and b are the attracting and repelling points of (k_i). As $a \neq b$, the terms of this sequence are eventually loxodromic. Since h_j is conjugate to k_i, this implies that h_j is in fact loxodromic.

On the other hand, suppose that for each j, there are finitely many distinct conjugates $g_i h_j g_i^{-1}$. We may thus pass to a subsequence of (g_i) such that for each $j \in \mathbb{N}$, there is $i(j) \in \mathbb{N}$, such that for $i \geq i(j)$, the sequence $(g_i h_j g_i^{-1})$ is constant. Define $\varphi \colon H \to G$ as $\varphi(h_j) = g_{i(j)} h_j g_{i(j)}^{-1}$, where $H = \{h_j \mid j \in \mathbb{N}\}$. It is straightforward to check that φ is injective.

Now let $j \in \mathbb{N}$. Observe that for $i \geq i(j)$, we have $\varphi(h_j)p_i = g_i p$, so that $\varphi(h_j)p_i \to b$. Similarly, $\varphi(h_j)q_i \to a$. Hence both a and b are fixed by each $\varphi(h_j)$. There is thus a convergence subsequence of $(\varphi(h_j))$ with attracting and repelling points $a \neq b$. This, in turn, shows that the elements $\varphi(h_j)$ are eventually loxodromic. Now h_j is conjugate to $\varphi(h_j)$, this shows h_j is loxodromic for sufficiently large j. $\square$

2.4. The space of triples. Starting with the boundary of a hyperbolic space, then we can try to reconstruct the space by considering triples of points in the boundary, viewed as vertices of ideal triangles. The subspace we get by taking 'centres' of these triangles is roughly well-defined. This is captured in the following exercise.

Exercise 2.27. Let X be a proper δ-hyperbolic metric space. Call a point $x \in X$ a *centroid* for an ideal triangle T if x is a distance of at most 10δ from each side of T. Show that every ideal triangle T has at least one centroid, and the distance between any two centroids is bounded by a constant depending only on δ. (Hint: approximate T with finite geodesic triangles, for which centroids are easy to find.)

In fact, when a group acts cocompactly on a hyperbolic space, the space of centroids is more or less the entire original space. We can view this as a sort of strong visibility property.

EXERCISE 2.28. Let X be a δ-hyperbolic metric space with a non-elementary, cocompact group action by isometries. Then there is some constant $K \geq 0$ such that every point of X is a distance of at most K from a centroid of an ideal triangle in X.

Mimicking the above, we can reconstruct a sort of abstract model for the 'interior' of an arbitrary compactum.

DEFINITION 2.29 (Space of triples). Let M be a topological space. Write $\Theta_0(M) = M^3 - \{(a,b,c) \mid \#\{a,b,c\} < 3\}$ for the space of distinct ordered triples of M, equipped with the product topology. The *space of triples* of M is the space $\Theta(M)$, obtained as the quotient of $\Theta_0(M)$ by the permutation action of the symmetric group on triples.

REMARK 2.30. When M is compact and metrisable, $\Theta(M)$ is locally compact and metrisable. If a group acts on M, then there is an obvious induced action on $\Theta(M)$.

We will, for the sake of convenience, largely ignore the above formalism and refer to elements of $\Theta(M)$ as three-element subsets of M. Convergence groups and their dynamical properties can be reformulated in terms of the topology of the action on the space of triples. Of course, when M has fewer than three points, $\Theta(M)$ is empty, so we usually discard this case.

To build a dictionary between actions on M and $\Theta(M)$, we will need to translate the proper discontinuity condition back into information about sequences.

LEMMA 2.31. *Suppose the action of G on $\Theta(M)$ is properly discontinuous, and $x_n \to x, y_n \to y, z_n \to z$, with $\{x,y,z\} \in \Theta(M)$. If (g_n) is a sequence of distinct group elements and $(g_n x_n), (g_n y_n)$, and $(g_n z_n)$ converge in M, at least two have a common limit point.*

PROOF. Suppose otherwise, so $g_n x_n \to x', g_n y_n \to y', g_n z_n \to z'$ with $\{x',y',z'\} \in \Theta(M)$. It follows that the sequence $\{g_n x_n, g_n y_n, g_n z_n\} \to \{x',y',z'\}$ converges in $\Theta(M)$. As $\Theta(M)$ is locally compact, $\{x,y,z\}$ has a compact neighbourhood K. But then $g_m^{-1} g_n K \cap K$ is non-empty for all sufficiently large values of m and n, contradicting proper discontinuity. □

LEMMA 2.32. *Suppose the action of G on $\Theta(M)$ is properly discontinuous, and let $x_n \to x, y_n \to y, z_n \to z$ be such that $\{x,y,z\} \in \Theta(M)$. If (g_n) is a sequence in G and $a \neq b \in M$ are such that $g_n x_n \to a$, $g_n y_n \to a$, and $g_n z_n \to b$, then (g_n) has a convergence subsequence with attracting point a and repelling point z.*

PROOF. Let $p \neq \{x,y,z\}$ and pass to a subsequence for which $(g_n p)$ and $(g_n^{-1} p)$ converge. By Lemma 2.31 applied to $(x_n), (z_n)$, and (p), either $g_n p \to a$ or $g_n p \to b$ after passing to a further subsequence. Suppose the latter is true, and pick some $c \neq a, b$. Pass again to a subsequence for which $w_n = g_n^{-1} c$ converges to $w \in M$. Now either $\{x,z,w\}$ or $\{y,p,w\}$ is a distinct triple: in each case apply Lemma 2.31 to $(x_n), (z_n)$, and (w_n) or $(y_n), (p)$, and (w_n). In either case we get that $(g_n w_n)$ converges to a or

b, but $g_n w_n = c \neq a, b$, a contradiction. Hence g_n converges pointwise to the constant function at a on $M - \{z\}$.

We must prove that this pointwise convergence is in fact uniform on compact subsets. Let $K \subseteq M - \{z\}$ be a compact subset and $U \subseteq M$ an open neighbourhood of a. Suppose that the convergence is not uniform, so that there is an infinite sequence (w_n) in K such that $g_n w_n \notin U$, After passing to a subsequence of (w_n), we may assume $w_n \to w \in K$. We may pass to a further subsequence for which $(g_n w_n)$ converges.

Since $z \notin K$ and K is closed, $w \neq z$. This implies that $g_n w \to a$ by pointwise convergence, and hence that there is a neighbourhood V of w with $b \notin g_n V$ for sufficiently large n. Thus we can apply Lemma 2.31 to $(x_n), (z_n)$, and (w_n) to see that that $g_n w_n \to a$ or $g_n w_n \to b$. In the former case, the fact that $g_n w_n \notin U$ immediately gives a contradiction, while in the latter case, we also obtain a contradiction from the fact that $w_n \in V$ for all sufficiently large n. □

THEOREM 2.33. *Let M be a compact metrisable space with at least three points, G a group acting by homeomorphisms on M. Then G is a convergence group on M if and only if the induced action on $\Theta(M)$ is properly discontinuous.*

PROOF. We first prove the forward direction: suppose G is a convergence group on M. Let $K \subseteq \Theta(M)$ be a compact subset and (g_i) is an infinite sequence of elements with $g_i K \cap K \neq \emptyset$. That is, there is a sequence $(\{x_i, y_i, z_i\})$ in K such that $\{g_i x_i, g_i y_i, g_i z_i\} \in K$ for each i also. Passing to a subsequence, we may assume that $\{g_i x_i, g_i y_i, g_i z_i\}$ converges to a point $\{x', y', z'\}$ in K, by compactness.

Now, as G is a convergence group, there is a convergence subsequence (g_{n_i}) with attracting point a and repelling point b. Without loss of generality, we may suppose that x_i and y_i are not equal to b. It follows that $x' = \lim g_i x_i = a$ and $y' = \lim g_i y_i = a$, contradicting the fact that x' and y' are distinct. Thus no such infinite sequence exists.

For the converse, suppose that G acts on $\Theta(M)$ properly discontinuously. Let (g_i) be an infinite sequence of distinct elements of G. Let $\{x, y, z\} \in \Theta(M)$. By Lemma 2.31, there is some subsequence (g_{n_i}) and points a and b in M such that $g_{n_i} x \to a$, $g_{n_i} y \to a$, and $g_{n_i} z \to b$ (after possibly relabelling). If $a \neq b$, then Lemma 2.32 completes the proof, so suppose otherwise, that $a = b$. Pick some $c \neq a$ and let $w_i = g_{n_i}^{-1} c$. Passing to a further subsequence we can assume $w_i \to w$ and, after possibly relabelling, $w \neq x, y$. But then $g_{n_i} x, g_{n_i} y \to a$ but $g_{n_i} w_i = c$. Again (g_{n_i}) has a convergence subsequence by Lemma 2.32. □

As a basic consequence of this characterisation, we can deduce that the only uncountable convergence groups arise in the trivial case of trivial convergence actions. One should view this as a manifestation of the essentially discrete nature of these groups. Note that this fact is certainly not obvious from the dynamical characterisation!

COROLLARY 2.34. *If G is a non-elementary convergence group, then G is countable.*

PROOF. Let G be a non-elementary convergence group on compact metrisable space M. By definition $\Lambda G \subseteq M$ has at least three points. Then by Proposition 2.33, G acts properly discontinuously on $\Theta(M)$, which is locally compact and metrisable. In particular, $\Theta(M)$ is locally compact, Hausdorff, and second countable. Since the action is properly discontinuous and the space is locally compact, the stabiliser of any point is finite. Moreover, the orbit of any point is discrete and second countable, and thus countable. Thus the quotient of G by the kernel of its action on $\Theta(M)$ is countable. The result now follows form the fact that the action of G on M has finite kernel, as in Remark 2.6. $\square$

As one might expect, more or less any dynamical criterion on a convergence group can be rephrased in terms of more topological criterion on the space of triples. For instance, being a uniform convergence group has the characterisation below, due to Tukia; the proof is somewhat long and we will not present it here.

THEOREM 2.35. *Let G be a convergence group on compact metrisable space M. Then G is uniform if and only if the action of $\Theta(M)$ is cocompact.*

The latter condition is frequently given as the definition of a uniform convergence group, as the notion is based on the behaviour of cocompact subgroups of $\mathrm{Isom}(\mathbb{H}^n)$. We will see soon that, more generally, hyperbolic groups are uniform convergence groups on the boundaries of spaces they act on geometrically.

3. Convergence groups and hyperbolic spaces

We finally turn the machinery developed in the previous section on groups acting on hyperbolic spaces. The key observation linking the two theories is the following.

THEOREM 3.1. *Let G be a group acting properly discontinuously by isometries on a proper hyperbolic metric space X. Then the induced action of G on ∂X is a convergence action.*

PROOF. Let X be a δ-hyperbolic metric space with properly discontinuous isometric G-action, and fix a basepoint $x \in X$. If G is finite there is nothing to prove, so suppose otherwise. Take (g_i) an infinite sequence of distinct elements of G. As the action is proper, the sequence $(g_i x)$ is unbounded in X. After possibly passing to a subsequence, there is some point $a \in \partial X$ such that $g_i x \to a$, since $X \cup \partial X$ is compact.

If for all sequences (b_i) in ∂X, we have $g_i b_i \to a$, then (g_i) is straightforwardly a convergence sequence with attracting and repelling point a, and so we are done. Otherwise, there is a sequence (b_i) in ∂X and $c \in \partial X$ such that $g_{n_i} b_i \to c$ with $c \neq a$, for some subsequence (g_{n_i}). By compactness of ∂X, we may pass to a further subsequence for which b_i converges to some point $b \in \partial X$. The claim is that (g_{n_i}) is a convergence sequence with attracting point a and repelling point b.

We need to show that for any compact subset $K \subseteq \partial X - \{b\}$ and any neighbourhood U of a, that $g_{n_i} K \subseteq U$ for sufficiently large i. Let K be such a subset. As ∂X is Hausdorff, K is closed and, in particular, does not intersect every neighbourhood of b. Without loss of generality, then, we may take K to be the complement of $U(b, r)$ for some $r \geq 0$. Suppose $p \notin U(b, r)$, so that $\langle p, b \rangle_x \leq r$. Note that if $q_i \to q$, then for any $\varepsilon > 0$ we have the containment $U(q_i, r + \varepsilon) \subseteq U(q, r)$ for sufficiently large i. Thus there is a sequence (y_i) in X asymptotic to p with $\langle y_i, b_i \rangle_x \leq 2r$ for all i.

Let $r \geq 0$ be large enough that $U \cap U(c, r) = \emptyset$, which exists since ∂X is normal Hausdorff. The observation above implies that $U(g_{n_i} b_i, 2r) \subseteq U(c, r)$ for sufficiently large i. Thus $\langle g_{n_i} x, g_{n_i} b_i \rangle_x \leq 2r$ for large i. It follows that, for i large enough, there is a point z_i on a geodesic $[g_{n_i} b_i, g_{n_i} x]$ with $\mathrm{d}(z_i, x) \leq 2r + 2\delta$.

Now we have

$$\langle g_{n_i} y_j, z_i \rangle_{g_{n_i} x} \leq \langle g_{n_i} y_j, g_{n_i} b_i \rangle_{g_{n_i} x} = \langle y_j, b_i \rangle_x \leq 2r,$$

where the first inequality comes from the fact that z_i is on the geodesic $[g_{n_i} b_i, g_{n_i} x]$, the equality from the fact that G acts by isometries, and the last inequality from the construction of (y_i).

Combining these with the triangle inequality implies that $\langle g_{n_i} y_j, x \rangle_{g_{n_i} x} \leq 6r + 4\delta$. As $\mathrm{d}(g_{n_i} y_j, x)$ is unbounded while $\langle g_{n_i} y_j, x \rangle_{g_{n_i} x}$ is bounded, we must have that $\langle g_{n_i} y_j, g_{n_i} x \rangle_x$ is unbounded in i. That is, for any $r' \geq 0$, there is i such that $\langle g_{n_i} y_j, g_{n_i} x \rangle_x \geq r'$. Being that $(g_{n_i} y_j)_{j \in \mathbb{N}}$ is asymptotic to $g_{n_i} p$, this means $g_{n_i} p \in U(a, r')$. Choosing $r' \geq 0$ large enough so that $U(a, r') \subseteq U$ thus completes the proof. □

REMARK 3.2. It is open whether the converse of the above theorem holds. That is, given a convergence group action of G on compact metrisable M, whether there is a proper hyperbolic space X admitting a properly discontinuous action by isometries such that ∂X is G-equivariantly homeomorphic to M.

It is true, by work of Sun, that convergence groups admit reasonably nice actions on hyperbolic spaces. However, the boundaries of these spaces do not recover the original compactum, and the actions are generally not properly discontinuous.

The classification of convergence group elements gives a somewhat more straightforward proof of the classification of isometries of a proper hyperbolic space.

THEOREM 3.3. *Every isometry of a proper hyperbolic space is elliptic, parabolic, or loxodromic.*

PROOF. Let X be a proper hyperbolic metric space, $g \in \mathrm{Isom}(X)$. If g does not have bounded orbits, then it must be an element of infinite order, and so $\langle g \rangle$ acts properly discontinuously on X. Hence $\langle g \rangle$ is a convergence group on ∂X. Every infinite order element of $\langle g \rangle$ is parabolic or loxodromic in the sense of a convergence group on ∂X, which coincides with the definitions for isometries of a hyperbolic metric space. □

Of course, a hyperbolic group acts as a convergence group on ∂G, but more is true. We first need the following observation.

EXERCISE 3.4. Let X be a proper hyperbolic metric space admitting a cocompact group action by isometries. If X is unbounded, then ∂X contains at least two points.

(Hint: use the fact that there is no non-trivial group action on the ray $[0, \infty)$ by isometries.)

THEOREM 3.5. *Let G be a group acting geometrically on a proper hyperbolic metric space X. Then G is a uniform convergence group on ∂X.*

PROOF (SKETCH): Let X be a δ-hyperbolic metric space that X acts on geometrically. In the former case ∂X is empty and there is nothing to prove. Suppose that G is infinite, then. We know that G acts as a convergence group on ∂X, since the action of G on X is properly discontinuous. We need to show that every point of G is a conical limit point.

Let $p \in \partial X$ be an arbitrary point. By Exercise 3.4, ∂X contains at least two points, so let $a \in \partial X - \{p\}$. Fix some $x \in X$ as a basepoint and let $B \subseteq X$ be a compact set with $x \in B$ and $G \cdot B = X$. Let ℓ be a geodesic line through x with endpoints p and a. Take (g_i) to be the sequence of elements of G for which $g_i B$ meets ℓ.

As ∂X is compact, we may pass to a subsequence of (g_i) such that $g_i p$ converges to a point $b \in \partial X$. We claim that the sequence (g_i) and the points a and b serve as witnesses to the fact that p is a conical limit point. For every $i \geq 0$, picking $y \in \ell$ far enough from x in the direction of p ensures that x is between $g_i x$ and $g_i y$. That is to say, there is a sequence $y_j \to p$ such that $\langle g_i x, g_i y_j \rangle_x$ is uniformly bounded for $j \geq i$. Since $g_i y_j \to g_i p$ as $j \to \infty$, there is a neighbourhood of a excluding $g_i p$ for all i. It follows that $b = \lim g_i p \neq a$.

We will take for granted that $g_i a \to a$, as this is easier to prove. Let $q \neq p, a$, and let $r = \langle a, q \rangle_x < \infty$. It follows from hyperbolicity that there is $z \in [a, q]$ such that $\mathrm{d}(x, z) \leq r + 2\delta$. But then $z_i = g_i z$ is a point on a geodesic $[g_i a, g_i q]$ with $\mathrm{d}(g_i x, g_i z) \leq r + 2\delta$. As $g_i x \in g_i B$ is uniformly close to the geodesic ℓ, it follows that $\langle g_i x, g_i q \rangle_x$ is roughly equal to $\mathrm{d}(x, g_i x)$. This latter quantity tends to infinity as $i \to \infty$, so that $g_i q$ lies in any neighbourhood of $a = \lim g_i x$ for sufficiently large i. That is, $g_i q \to a$ as required. □

Definition 3.6. Let G be a hyperbolic group. Suppose that G acts geometrically on a proper hyperbolic space X. The *boundary* ∂G of G is the space ∂X.

The boundary exists for any hyperbolic group: one can take the Cayley graph with respect to any finite generating set as the space X. Moreover, ∂G is well-defined up to homeomorphism for a given hyperbolic group G. Indeed, the Milnor–Schwarz lemma tells us that if X and Y are two proper hyperbolic spaces admitting geometric actions by G with X hyperbolic, there is a G-invariant quasi-isometry $X \to Y$. Since quasi-isometries of proper hyperbolic spaces induce homeomorphisms of the boundary, there is a G-equivariant homeomorphism $\partial X \to \partial Y$ whenever this is the case.

The previous theorem does have a converse, due to Bowditch, so that we have a completely dynamical reformulation of hyperbolic groups. However, it is much more difficult to prove, so we only give a very rough sketch here.

Theorem 3.7. *Let G be a uniform convergence group of compact metrisable space M. Then G is a hyperbolic group and M is G-equivariantly homeomorphic to ∂G.*

Proof idea: Recall that we mentioned that a hyperbolic metric space X can be recovered up to quasi-isometry from a quasi-conformal structure on its boundary ∂X. The idea is as follows: analysing the action of G on M, we can equip M with a 'system of G-invariant annuli', that encode the structure of the group action. This is the key place where the uniformity of the convergence action is used, which ensures that such a system exists around each point.

A system of annuli is essentially a quasi-conformal structure on M, and we can use it to construct a version of a cross-ratio on M. Now, a cross-ratio on M induces a G-invariant *quasi-metric* on the space of triples $\Theta(M)$, which one can show is hyperbolic. One can then upgrade this quasi-metric to an actual G-invariant metric d on $\Theta(M)$. Since G is a uniform convergence group on M, it acts geometrically on the hyperbolic metric space $(\Theta(M), \mathrm{d})$. Thus $\partial\Theta(M) \cong M$ is G-equivariantly homeomorphic to ∂G by the observation before the theorem. □

Using the machinery we developed in the previous section, we are now able to deduce many strong facts about hyperbolic groups and their subgroups. First, we see that we have a more straightforward proof of the fact that every infinite order element of a hyperbolic group is loxodromic.

THEOREM 3.8. *Let G be a hyperbolic group. Then every infinite order element of G is loxodromic.*

PROOF. By Theorem 3.5, G acts as a uniform convergence group on ∂G. By Proposition 2.25, G contains no parabolic elements. But every infinite order element of G is parabolic or loxodromic by Lemma 2.9. □

We can also prove that hyperbolic groups contain no infinite torsion subgroups, as promised earlier. We state a simple lemma beforehand.

LEMMA 3.9. *Let H be an infinite subgroup of a uniform convergence group G. Then ΛH contains at least two points.*

PROOF. As H is infinite, ΛH is non-empty. Suppose that ΛH contains only a single point p. Necessarily, H fixes p, which is a conical limit point of G. But any infinite set fixing a conical limit point contains a loxodromic element. It follows that ΛH contains at least two points, the poles of this loxodromic. □

THEOREM 3.10. *Let G be a hyperbolic group. Then G contains no infinite torsion subgroups.*

PROOF. Let G be a hyperbolic group, so G is a uniform convergence group on ∂G. Let $H \leqslant G$ be an infinite subgroup, so that $\Lambda H \subseteq \partial G$ is non-empty. By Lemma 3.9, ΛH contains at least two points. We saw that H contains a loxodromic element (which generates a finite index subgroup of H) when ΛH contains two points. Now if ΛH contains at least three points, then H is non-elementary and so contains a non-abelian free subgroup by the Tits alternative for convergence groups. □

Another consequence of the above argument is a strong version of Tits' subgroup dichotomy for hyperbolic groups. Note that there are examples of non-linear hyperbolic groups, so the setting is really different to the original theorem of Tits.

THEOREM 3.11 (Tits' alternative). *Let G be a hyperbolic group, $H \leqslant G$ a subgroup. Then H is either virtually cyclic, or H contains a non-abelian free subgroup.*

We saw in Chapter 2 that any compact metrisable space arises as the boundary of a proper hyperbolic metric space. In contrast, we have already seen that even the cardinalities of boundaries of hyperbolic groups are quite restricted, as the group action requires the boundary to have some sort of uniform symmetry.

The topology of boundaries is, indeed, also very constrained, to the extent that in (very) low dimensions, one can in fact classify exactly the spaces that arise as boundaries. A majority of the work lies in some deep results about the topology of the boundary of hyperbolic groups: they are always locally connected when they are connected, and they contain no global cut points. The classification of one-dimensional continua then gives the following.

THEOREM 3.12. *Let G be a nonelementary hyperbolic group. If ∂G has dimension 0, then ∂G is a Cantor space. If ∂G is connected and has dimension 1, it is either a circle S^1, a Sierpiński carpet, or a Menger sponge.*

Note that these are all very regular spaces. Indeed, the circle is the only compact manifold in dimension one, the Sierpiński carpet is the universal plane curve, and the Menger sponge is the universal curve.

More can be said about the groups with these low-dimensional boundaries. Free groups act geometrically on trees whose boundaries are Cantor sets. Deep structural results on splittings of groups of Stallings and Dunwoody actually imply the following.

THEOREM 3.13. *Let G be a hyperbolic group with ∂G a Cantor set. Then G contains a finite index non-abelian free subgroup.*

Similarly, we have seen that fundamental groups of hyperbolic surfaces are natural examples of hyperbolic groups with circle boundary. In fact, these turn out to be the only such groups, up to finite index.

THEOREM 3.14. *Let G be a hyperbolic group with $\partial G \cong S^1$. Then G acts geometrically on the hyperbolic plane $\mathbb{H}^2$.*

This theorem combines the work of many authors, most notably Tukia and Gabai, and was also independetly proven by Casson and Jungreis. In reality, the theorem is a little stronger: it says that any convergence group action on the circle S^1 is conjugate in Homeo(S^1) to one induced by a properly discontinuous action by isometries on the hyperbolic plane $\mathbb{H}^2$. Again, the proof is long and complicated, but we give a very brief overview.

PROOF IDEA: Let G be a convergence group on S^1. We extend this action to a convergence group action on the disc D^2 by cutting up the disc using axes of loxodromic elements. If G contains a simple axis (that is, an axis disjoint from all its conjugates, then these axes cut the disc up into disjoint pieces that all meet the boundary circle, and one can extend the action in a canonical way to the interior of the disc. Tukia showed that this covers many cases, and the work of Gabai covers the rest of the cases. Once one has this, we can pick a G-equivariant triangulation of the disc D^2, which we can modify to resemble $\mathbb{H}^2$ by applying a conformal transformation of the disc. This is tantamount to conjugating the original action by a homeomorphism of the circle. $\square$

Embedding a Cantor set in a circle, one can actually deduce Theorem 3.13 from the above. Indeed, the action of the group on the Cantor set can be extended to an action on the circle, and the original group will act cocompactly on the convex hull of this embedded Cantor set. The quotient is (up to passing to a finite cover) a surface with boundary, which retracts onto a graph. Thus these groups have finite index free subgroups.

Moving one dimension up, the analysis becomes apparently much more difficult. For instance, the following still remains open after almost forty years.

CONJECTURE 3.15 (Cannon conjecture). *Let G be a hyperbolic group with $\partial G \cong S^2$. Then G acts geometrically on $\mathbb{H}^3$.*

Similarly to how the classification of groups with S^1 boundary shows that groups with Cantor set boundary acts geometrically on a convex subset of $\mathbb{H}^2$, a resolution to the above conjecture would allow a description of the groups with Sierpiński carpet boundary as those acting geometrically on a convex subset of $\mathbb{H}^3$. Indeed, a Sierpiński carpet boundary embeds into a sphere, with the holes as round circles. There are finitely many orbits of these circles under the action of the group, and their stabilisers are well-behaved hyperbolic subgroups with circle boundary (and so, they act geometrically on $\mathbb{H}^2$). 'Doubling' the group along representatives of these finitely many conjugacy classes of stabiliser subgroups yields a hyperbolic group containing the original, and whose boundary is the entire sphere. Then, if the Cannon conjecture were true, we could obtain an action of the original group on a convex subset of $\mathbb{H}^3$ from this.

In even higher dimensions, the versions of this geometric conjecture are also open. However, in high enough dimensions, techniques from surgery theory allow a resolution to a sort of topological analogue.

THEOREM 3.16 (Bartels–Lück–Weinberger). *Let G be a torsion-free hyperbolic group with $\partial G \cong S^{n-1}$, where $n \geq 6$. Then $G \cong \pi_1 M$, where M is an aspherical n-manifold with $\widetilde{M} \cong \mathbb{R}^n$.*

To conclude our discussion on boundaries, we state another result that emphasises the regularity of boundaries of hyperbolic groups. It tells us that any boundary that contains a subset homeomorphic to $\mathbb{R}^n$ is actually an n-sphere. This rules out any non-sphere manifolds arising as boundaries of hyperbolic groups. Note that the same argument actually works for the limit set of any non-elementary convergence group.

THEOREM 3.17. *Let G be a hyperbolic group and suppose that ∂G contains a subset homeomorphic to an open subset of $\mathbb{R}^n$. Then ∂G is homeomorphic to the n-sphere S^n.*

PROOF. The idea is to use the convergence property and a well-chosen loxodromic to show that ∂G is the union of two open n-balls glued along their boundary. We will need the generalised Schoenflies theorem, which states that a bicollared topologically embedded sphere in $\mathbb{R}^n$ separates it into two components, one bounded and one unbounded. This is the higher dimensional version of the Jordan curve theorem.

Let $U \subseteq \partial G$ be a subset homeomorphic to $\mathbb{R}^n$. By the density of loxodromic fixed point pairs, there is a loxodromic element $g \in G$ with $P_g, N_g \in U$. Let U_+ and U_- disjoint open neighbourhoods of P_g and N_g respectively, and let $f_+ \colon \mathbb{R}^n \to U_+$ and $f_- \colon \mathbb{R}^n \to U_-$ be homeomorphisms such that $f_+(0) = P_g$ and $f_-(0) = N_g$. Let B_1 and B_2 be the open balls of radius 1 and 2 in $\mathbb{R}^n$ respectively. We define $V_+ = f_+(B_2)$ and $V_- = f_-(B_1)$, which are neighbourhoods of P_g and N_g respectively. Since (g^m) is a convergence sequence, there is $m \geq 0$ such that

$$g^m(\partial G - V_-) \subseteq V_+.$$

Let $S = f_-(\partial B_1)$, so that S is a sphere that is the topological boundary of V_- in ∂G. There is some $\varepsilon > 0$ such that $S \times [-\varepsilon, \varepsilon]$ embeds in ∂G also, so S is a bicollared sphere. Of course, $S \subseteq \partial G - V_-$, so that $g^m S \subseteq V_+$. Now $S' = f_+^{-1}(g^m S) \subseteq B_2$ is a bicollared sphere in $\mathbb{R}^n$, and so by the generalised Schoenflies theorem separates $\mathbb{R}^n$ into two components, one bounded and homeomorphic to an open n-ball D^n, the other unbounded and homeomorphic to the complement of a closed n-ball in $\mathbb{R}^n$. Since $\partial G - V_-$ is compact (as a closed set in a compact Hausdorff space) whose topological boundary is S, its image under $f_+^{-1} g^m$ is exactly the closure of the bounded component of $\mathbb{R}^n - S'$.

Therefore $\partial G - V_-$ is homeomorphic to a closed n-ball. By construction, V_- is homeomorphic to an open n-ball. The two sets share a topological boundary, which is the $(n-1)$-sphere S. Thus $\partial G = D^n \cup_{S^{n-1}} D_n \cong S^n$, as required. □

4. Algorithms and computability

Finite presentations are inherently a tool for practical computation and combinatorial manipulation. Accordingly, the most basic questions one can ask about them are related to computability. The first three such questions were originally posed by Dehn in the early twentieth century.

We will not concern ourselves with a precise notion of computability here: consider the truth of a statement *decidable* if there is an algorithm (that is, a sequence of operations) one can perform which, after finitely many steps, will return that the statement is either true or false. The *word* and *conjugacy problems* ask whether presentations can allow us to meaningfully distinguish elements of the groups they define.

Definition 4.1. Let $P = \langle S \mid R \rangle$ be a finite presentation. We say the *word problem* is solvable in P if, given any two words w and v in S, the statement that w and v represent the same element is decidable. Similarly, we say that the *conjugacy problem* is solvable in P if the statement that w and v represent conjugate elements is always decidable.

It is straightforward, by applying Tietze transformations, to see that having solvable word or conjugacy problem is really a property of the group that a given finite presentation defines, and is independent of the choice of finite presentation for that group. Hence we may rightly call solvability of these problems group properties. Of course, the word problem is a special case of the conjugacy problem, since the conjugacy class of the identity contains only the identity.

Remark 4.2. The word problem is always *semi-decidable*: there is an algorithm such that, if w and v in fact *do* represent the same element, will eventually halt and return true. Namely, one can in a naïve way enumerate all words representing the identity in a finite presentation, and observe that if w and v represent the same element, the word $w^{-1}v$ will appear somewhere in this list. Of course, this list is infinite, so the algorithm will never terminate if w and v represent different elements.

Exercise 4.3. Solve the word problem for finitely presented simple groups.

Typically harder, though not as obviously so, is the problem of distinguishing one finite presentation from another.

Definition 4.4. Let $\mathcal{C}$ be a class of groups. We say the *isomorphism problem* is solvable in $\mathcal{C}$ if, for any two finite presentations P and Q of groups in $\mathcal{C}$, the statement that P and Q are presentations of isomorphic groups is decidable.

If one could solve the isomorphism problem over the class of all finitely presented groups, then certainly one has a solution to the conjugacy (and hence, also word) problem, as one can decide whether the presentation obtained by adjoining the desired conjugacy relation gives a different group or not.

A landmark result, first achieved by Novikov and, independently, Boone in the 1950s, provides a decidedly negative answer to all of these problems.

THEOREM 4.5 (Novikov–Boone). *There is a finitely presented group with unsolvable word problem.*

The proof of the above is quite difficult, and has deep ties to mathematical logic. Before moving back to hyperbolic groups, we mention a particularly striking result obtained independently by Adian and Rabin, around the same time as the Boone–Novikov result above. They showed that determining whether finite presentation has essentially any interesting property is undecidable.

We say that a group property P is a *Markov property* if there exists a finitely presented group with P, and also a finitely presented group that is not a subgroup of any finitely presented group with P. Among such properties are being finite, having solvable word problem, and being hyperbolic.

THEOREM 4.6 (Adian–Rabin). *Let P be a Markov property. It is undecidable whether any given finite presentation defines a group that has P.*

As a contrast to the rather unpleasant situation one finds oneself in for finitely presented groups in general, hyperbolic groups have excellent computability properties. In fact, all three of the above problems are solvable within the class of hyperbolic groups. In the remainder of this section, we present a solution to the word problem. The conjugacy problem also admits a solution that is not much more difficult, while the isomorphism problem requires a large number of advanced tools to solve.

DEFINITION 4.7 (Dehn presentation). A finite presentation P is called a *Dehn presentation* if any word in the presentation which represents the identity contains more than half of a relator.

Dehn presentations are so named because they exhibit the key property possessed by the standard presentation of higher genus surface groups which was used by Dehn to solve the world problem in such groups. Indeed, having a Dehn presentation yields a very easy solution to the word problem.

LEMMA 4.8. *The word problem is solvable in a group with a Dehn presentation.*

PROOF. Let $\langle S \mid R \rangle$ be a Dehn presentation for a group G, and let w be a word in S. We may reduce the word w as follows: if there is a relator $r = r_1 r_2 \in R$ with $\ell(r_1) < \ell(r_2)$ and w contains r_2 as a subword, then let w' be the word obtained from w by replacing this instance of r_2 with r_1^{-1}. By construction $\ell(w') < \ell(w)$, and w' represents the same element of G as w.

Applying such a reduction finitely many times, we obtain a word w'' representing the same element of G as w, none of whose subwords are more than half of a relator. Since the presentation was a Dehn presentation, w'' represents the identity if and only if it is the empty word, deciding the problem. □

The solution to the word problem in hyperbolic groups is a consequence of the fact that every hyperbolic group has a Dehn presentation. This not only shows that hyperbolic groups are finitely presentable, but finitely presentable in a very effective way.

THEOREM 4.9. *Let G be a hyperbolic group. Then G has a Dehn presentation.*

PROOF. Let S be a finite generating set for G, so that $X = \Gamma(G, S)$ is δ-hyperbolic. Let $k \geq 0, \lambda \geq 1, c \geq 0$ be constants such that every k-local geodesic is (λ, c)-quasigeodesic, which exist since X is hyperbolic. Let R be the set of cyclically reduced words in S with length at most $\max\{2k, \lambda c\}$. We will show that $\langle S \,|R\rangle$ is a Dehn presentation for G.

Suppose that w is a word representing the identity in G. We may cyclically reduce w, if it is not already cyclically reduced. Let p be the loop in X based at the identity, obtained by following the edges corresponding to the letters of w. Consider first the case that that p is a k-local geodesic. Then p is a (λ, c)-quasigeodesic with length $\ell(w)$. Since the distance between the endpoints of p is zero, it follows that $\ell(w) \leq \lambda c$. Hence $w \in R$.

Now if p is not a k-local geodesic, then w contains a minimal subword v of length at most k whose corresponding subpath q of p is not geodesic. Let q' be a geodesic in X with the same endpoints as q, and let u be the word corresponding to q'. Of course, q' also has length at most k, As v was minimal, q and q' have no overlap, so vu^{-1} is cyclically reduced. Moreover as the concatenation of q and q' is a loop in X, the word vu^{-1} represents the identity. Finally, uv^{-1} has length at most $\ell(q) + \ell(q') \leq 2k$, so $vu^{-1} \in R$. Moreover, $\ell(u) < \ell(v)$ since q was not a geodesic, so that w contains more than half of a word in R as required. Thus, if $\langle S \,|\, R\rangle$ is a presentation, it is a Dehn presentation.

To conclude, observe that if $w = w_1 v w_2$ and $r = vu^{-1} \in R$, we have

$$\begin{aligned} w &= w_1 v u^{-1} u w_2 \\ &= w_1 v u^{-1} w_1^{-1} w_1 u w_2 \\ &= (w_1 r w_1^{-1})(w_1 u w_2). \end{aligned}$$

If, further, $\ell(u) < \ell(v)$, then $w_1 u w_2$ is a strictly shorter word than w. By a finite induction, then, every word representing the identity in G may be written as a product of conjugates of elements of R. Hence $\langle S \,|\, R\rangle$ defines a genuine presentation of G. □

COROLLARY 4.10. *The word problem is solvable in hyperbolic groups.*

REMARK 4.11. As it turns out, having a Dehn presentation is equivalent to hyperbolicity, though we do not prove it here. Thus, computability is in some ways intrinsically tied to hyperbolicity.

EXERCISE 4.12. Solve the conjugacy problem in hyperbolic groups.

(Hint: Let w and v be cyclically reduced words, and suppose that they are conjugate by u, so $uwu^{-1} = v$. Take u to be such a conjugator with minimal length. If $\ell(u)$ is much larger than $\ell(w)$ and $\ell(v)$, the geodesic rectangle in a Cayley graph with sides labelled by w, u^{-1}, v, u is very long and thin. Use this to bound the length of such u, and hence effectively decide whether two words are conjugate.)

5. Small cancellation theory

One of the original motivations driving the development of the theory of hyperbolic groups was the so-called *small cancellation theory*, a collection of results and techniques that had been taking shape over several decades beforehand. Small cancellation theory is part of combinatorial group theory, the study of groups by their presentations. Though the ideas are somewhat geometric in nature, the mathematics is mostly combinatorial; hyperbolicity puts many of the results of the theory on a coherent geometric framework.

The main theme of small cancellation theory is the analysis of presentations where the relators do not have significant overlap (hence the 'small cancellation'). There are many different 'small cancellation conditions'; here, we state just one of the most common and important among them.

Definition 5.1 (Piece). Let $\langle S \mid R \rangle$ be a group presentation, and suppose that each relator $r \in R$ is cyclically reduced. Suppose further that R is *symmetrised*, so that R is closed under cyclic permutations and inverses of words.

A non-trivial word w in S is called *piece* of the presentation if w is a maximal common prefix for two distinct relators in R.

Definition 5.2 ($C'(\lambda)$ small cancellation condition). Let $\langle S \mid R \rangle$ be a group presentation, $\lambda > 0$. We say that the presentation satisfies the $C'(\lambda)$ *condition* if, whenever w is a piece of a relator $r \in R$, we have $\ell(w) < \lambda \ell(r)$.

Many (but far from all!) facts about groups satisfying the above classical small cancellation condition have been subsumed by hyperbolic groups.

Theorem 5.3. *Let $G = \langle S \mid R \rangle$ be a finitely presented group, with the presentation satisfying $C'(\lambda)$ for $\lambda \leq \frac{1}{6}$. Then G is a hyperbolic group.*

We do not give a proof of the above theorem, as it would be too much of a diversion. One should think of the $\frac{1}{6}$ appearing in the statement with regards to classical geometry. The tightest tiling of the Euclidean plane by regular polygons is that by triangles, with at most six triangles touching each point; if one wants to put more triangles in, one has to turn to the hyperbolic plane. Classical small cancellation theory inherently deals with planar diagrams and their geometry, so this link is in some senses very explicit.

We now turn to a remarkable construction due to Rips, which allows one to generate hyperbolic groups with many pathological properties.

Theorem 5.4 (Rips' construction). *Let Q be a finitely presented group. There is a hyperbolic group G with 2-generated normal subgroup $N = \langle x, y \rangle \lhd G$ such that $G/N \cong Q$.*

Proof. Let $\langle S \mid R \rangle$ be a finite presentation for Q, with finite sets $S = \{s_1, \dots, s_n\}$ and $R = \{r_1, \dots, r_m\}$. For simplicity, suppose that S is symmetric. We write $S' = S \cup \{x, y\}$, where x and y are two additional letters. For each $i = 1, \dots, n$, define the words in S':

$$t_{i,x} = s_i x s_i^{-1} x y^{a_i} x y^{a_i+1} \dots x y^{a'_i}$$
$$t_{i,y} = s_i y s_i^{-1} x y^{b_i} x y^{b_i+1} \dots x y^{b'_i},$$

and for $i = 1, \dots, m$:

$$r'_i = r_i x y^{c_i} x y^{c_i+1} \dots x y^{c'_i},$$

where $a_i < a'_i < b_i < b'_i < c_i < c'_i$. Now let

$$R' = \{r'_1, \dots, r'_m, t_{1,x}, \dots, t_{n,x}, t_{1,y}, \dots, t_{n,y}\}.$$

We define the group G via the presentation $\langle S' \mid R' \rangle$.

We first verify that G is hyperbolic. In fact, we will show that the presentation given is a $C'(1/6)$ presentation, and therefore hyperbolic. By the choice of the integers above, the noise words that suffix each relator in R' contain no pieces. Hence any piece of a relator in R' is a subword of some $r \in R$. Hence, choosing our integers such that $\min\{a_i' - a_i, b_i' - b_i\} \geq 100$ and $c_i' - c_i \geq 100 \max\{\ell(r) \mid r \in R\}$ ensures that the presentation is $C'(1/6)$.

Next, we show that $N = \langle x, y \rangle$ is a normal subgroup of G. Indeed, the relations $t_{i,x}$ and $t_{i,y}$ guarantee that the conjugate of x or y by any of the generators $s \in S$ lies in N. Since G is generated by S and $\{x, y\}$, it follows that N is normal in G. Lastly, G/N has the presentation $\langle S' \mid R' \cup \{x, y\} \rangle$. Applying Tietze transformations, one immediately sees that this recovers $Q = \langle S \mid R \rangle$, as required. $\square$

Let us give a couple of applications of the above to construct pathological subgroups of hyperbolic groups. We say that a finite presentation has *solvable membership problem* if there is an algorithm that can determine whether a given word represents an element of a given finitely generated subgroup. Again, of course, this is independent of the choice of finite presentation for a given group. The word problem is a special case of this.

Corollary 5.5. *There is a hyperbolic group with unsolvable membership problem.*

Proof. Let Q be a finitely presented group with unsolvable word problem. Applying the Rips construction to Q, we obtain a hyperbolic group G with 2-generated normal subgroup N such that $G/N \cong Q$. If the membership problem were solvable in G, then there would be an algorithm for deciding whether a word in a generating set for G represents an element of N. But this yields a solution to the word problem in Q, for an element in Q is non-trivial if and only if it has a lift in G that is not in N. $\square$

Corollary 5.6. *There is a hyperbolic group with a finitely generated subgroup that is not finitely presentable.*

Proof. Let Q be a finitely presented group containing a finitely generated but not finitely presentable subgroup P, for instance a free group of rank 2. Applying the Rips construction to Q, we obtain a hyperbolic group G with Q as a quotient and finitely generated kernel N. Then the preimage K of P in G under the quotient $G \to Q$ is generated by preimages of generators of P together with N. As N is finitely generated, K is also finitely generated. However, since N is normal, adding the generators of N as relators to any presentation of K yields a presentation of P. Hence K cannot be finitely presented, as P is not finitely presented. $\square$

Note that it follows from more advanced methods that N is not finitely presentable whenever Q is infinite: one can show that G has *cohomological dimension* 2, and among such groups any finitely presented normal subgroup is either finite, free, or has finite index in G. We will see the kernel of the Rips construction is infinite and never free, and if Q is infinite, it must have infinite index.

6. Quasiconvex subgroups

The Rips construction illustrates that general finitely generated subgroups of hyperbolic groups may be quite poorly behaved. We will touch on a class of subgroups whose intrinsic geometry somehow respects that of their ambient group, and are as a result much better behaved. Recall that a subspace of a metric space is *quasiconvex* if any geodesic with endpoints in the subspace is contained in a uniform neighbourhood of that subset.

DEFINITION 6.1. Let G be a group with finite generating set S. We say that $H \leqslant G$ is *quasiconvex* if there is $\sigma \geq 0$ such that H is σ-quasiconvex as a subspace of $\Gamma(G, S)$.

A priori quasiconvexity of a particular subgroup is dependent on the choice of generating set of the ambient group.

EXERCISE 6.2. Find a finitely generated group G and finitely generated subgroup $H \leqslant G$ such that H is quasiconvex with respect to one generating set but not another.

Note that we could equivalently define quasiconvexity in terms of actions: given G acting on X geometrically, we say $H \leqslant G$ is quasiconvex if there is a H-invariant subspace $Y \subseteq X$ such that H acts on Y geometrically. This definition is equivalent to the previous by the Milnor–Schwarz lemma. Similarly to how the previous definition depended on the choice of generating set, this one depends on the choice of action.

For a hyperbolic group, quasiconvex subgroups coincide exactly with the finitely generated quasi-isometrically embedded subgroups. As a consequence, being quasiconvex is independent of choice of generating set, as a change of finite generating set gives a quasi-isometry.

LEMMA 6.3. *Let G be a hyperbolic group with finite generating set S, and $H \leqslant G$ a subgroup with finite generating set T. Then H is quasiconvex in G if and only if the inclusion map $\iota \colon (H, \mathrm{d}_T) \to (G, \mathrm{d}_S)$ is a quasi-isometric embedding.*

PROOF. Fix finite generating sets S of G and T of H. Quasi-isometrically embedded subspaces of hyperbolic metric spaces are quasiconvex by Lemma 3.6, giving the backwards implication. Conversely, if (H, d_T) is σ-quasiconvex in $\Gamma(G, S)$ by Lemma3.4 implies that the inclusion $H \hookrightarrow G$ is quasi-isometry, with respect to the metric $\mathrm{d}_{H,2\sigma+1}$. After possibly enlarging T to T' include all of the elements finitely many elements $h \in H$ with $|h|_S \leq 2\sigma + 1$, the identity map on H is a quasi-isometry of these two metrics. Hence the inclusion map $(H, \mathrm{d}_{T'}) \to (G, \mathrm{d}_S)$ is a quasi-isometric embedding. Since all finite generating sets induce quasi-isometric embeddings, this proves the lemma. □

We also saw earlier in the course that quasi-isometrically embedded subsets of a hyperbolic space are again hyperbolic. The above thus yields:

COROLLARY 6.4. *Quasiconvex subgroups of hyperbolic groups are hyperbolic.*

EXERCISE 6.5. Show that every finitely generated subgroup of a free group of finite rank is quasiconvex.

EXERCISE 6.6. Show that being quasiconvex passes to finite index subgroups and overgroups.

An important feature of quasiconvexity is their intersection closure.

PROPOSITION 6.7. *Let G be a hyperbolic group. If H and K are quasiconvex subgroups of G, then so is $H \cap K$.*

PROOF. Let S be a finite generating set for G and $\sigma \geq 0$ be a quasiconvexity constant for H and K. We will show that for any $r \geq 0$ there is $R \geq 0$ such that

$$N_r(H) \cap N_r(K) \subseteq N_R(H \cap K).$$

Suppose otherwise, so that for each $n \in \mathbb{N}$ there is an element $x_n \in N_r(H) \cap N_r(K)$ with $\mathrm{d}_S(x_n, H \cap K) \geq n$. By definition, there are $y_n, z_n \in G$ for every n with the property that $|y_n|_S, |z_n|_S \leq r$, and $y_n x_n = h_n \in H$ and $z_n x_n = k_n \in K$. Rearranging, we have that $x_n = y_n^{-1} h_n = z_n^{-1} k_n$.

Since there are only finitely many $g \in G$ with $|g|_S \leq r$, we may pass to a subsequence for which $y_n = y$ and $z_n = z$ are constant. Hence we have that $h_n k_n^{-1} = yz^{-1}$ for all n In particular, $h_n k_n^{-1} = h_1 k_1^{-1}$, and so $h_1^{-1} h_n = k_1^{-1} k_n \in H \cap K$ for all n. But then $x_1^{-1} x_n = h_1^{-1} y y^{-1} h_n = h_1^{-1} h_n \in H \cap K$, so that $x_n \in x_1(H \cap K)$. This means that $\mathrm{d}_S(x_n, H \cap K) \leq |x_1|_S$, a contradiction for large enough n. This proves the claim.

Now the claim gives us $\Sigma \geq 0$ such that $N_\sigma(H) \cap N_\sigma(K) \subseteq N_\Sigma(H \cap K)$. That $H \cap K$ is Σ-quasiconvex as a subset of $\Gamma(G, S)$ then follows immediately from the fact that any geodesic with endpoints in $H \cap K$ lies in $N_\sigma(H)$ and $N_\sigma(K)$. □

Quasiconvexity also generally seems to be at odds with normality; we sketch a proof of the following.

PROPOSITION 6.8. *Let G be a hyperbolic group, $H \leqslant G$ a quasiconvex subgroup. If H is normal in G, then H is either finite or has finite index in G.*

PROOF (SKETCH): Suppose that H is infinite. Then ΛH is a closed non-empty subset of ∂G. As H is normal, for any $x \in \partial G$ and $g \in G$ we have $g \cdot Hx = H(gx)$, so that ΛH is G-invariant. But since the action of G on ∂G is minimal, it must be that $\partial G = \Lambda H$. Thus H has finite index in G. □

We note that the finite normal subgroups of a hyperbolic group G all arise in a somewhat trivial way.

LEMMA 6.9. *Let G be a hyperbolic group, and let $K \lhd G$ be the kernel of the action of G on ∂G. Then every finite normal subgroup of G is contained in K.*

PROOF. Let $F \lhd G$ be a finite normal subgroup. The quotient map $q \colon G \to G' = G/F$ is a quasi-isometry, so induces an equivariant homeomorphism of boundaries $\partial G \to \partial G'$. Now F acts trivially on $\partial G'$, so it must act trivially on ∂G. Hence $F \subseteq K$. □

EXERCISE 6.10. Let G be a hyperbolic group. Show that if $H \leqslant G$ is quasiconvex, then H has finite index in its centraliser $C_G(H)$.

7. Topological properties

Hyperbolic groups also have properties that make them particularly well-behaved from the perspective of algebraic topology. Let us give a bit of background to contextualise the upcoming results. For a topological space X, the (co)homology groups $H_i(X)$ and $H^i(X)$ are incredibly important algebraic invariants that carry a wealth of topological data. If X is, for instance, a finite simplicial or cellular complex, then these groups can be computed explicitly (at least, in theory). Part of the power of algebraic topology is that very many spaces of interest may be modelled by such finite complexes, and so one can apply all of the wonderful tools of algebraic topology to study these spaces.

In much a similar fashion, there is a theory of *group (co)homology*, which allows one to define the functors H_i and H^i on the category of groups. Analogously with the situation for spaces, the groups $H_i(G)$ and $H^i(G)$ may encode a great deal of algebraic information about the group G. As such, it is very useful to have finite models that allow us to compute these groups. For a (discrete) group G, a *classifying space* is a space BG for which $\pi_1(BG) = G$ and $\pi_n(BG) = 0$ for all $n \geq 2$. In the simplest case of (co)homology with integral coefficients, the groups $H_i(G)$ and $H^i(G)$ are equal to the groups $H_i(BG)$ and $H^i(BG)$ respectively, so BG serves as a model for G. In this section, we will show that torsion-free hyperbolic groups have finite classifying spaces.

We note that the name 'classifying space' refers to the fact that such a space BG classifies principal G-bundles in the sense that BG is the orbit space of a weakly contractible space EG by a free and transitive action, and any principal G-bundle $Y \to Z$ is a pullback of a map $Z \to BG$ over $EG \to BG$. That is to say, principal G-bundles over Z are in one-to-one correspondence with maps $Z \to BG$.

We will make use of the following construction, which turns metric spaces into simplicial complexes whose simplices in some sense coarsely approximate balls in the metric space.

Definition 7.1 (Rips complex). Let X be a metric space, $r \geq 0$. The *Rips complex* on X with parameter r is the complex $P_r(X)$ whose vertex set is X and with an n-simplex for every $(n+1)$-tuple $Y = \{x_0, \dots, x_n\}$ with $\mathrm{diam}(Y) \leq r$.

The Rips complex, taken with a suitably large parameter, is always contractible for a hyperbolic metric space. It will therefore serve as a candidate for the universal cover EG of our classifying space.

Proposition 7.2. *Let X be a δ-hyperbolic space, $X' \subseteq X$ a subspace with $X = N_1(X')$. If $r \geq 4\delta + 2$ then $P_r(X')$ is contractible.*

Proof. Let $r \geq 4\delta + 2$. By Whitehead's theorem, it suffices to show that the homotopy groups of $Y = P_r(X')$ are trivial. Pick a basepoint $x \in X'$ and suppose that $S^n \to Y$ is a continuous map of a sphere into Y based at x. Since S^n is compact, the image of this map lies in a finite subcomplex of Y. To show that the map is null-homotopic, we will show that every finite subcomplex of Y is contractible.

Let $L \subseteq Y$ be a finite subcomplex. We will homotope L to a strictly smaller complex l' by moving vertices of L towards the basepoint x. Repeating such a process finitely many times, L is homotopic to a finite subcomplex of Y in which every vertex is a distance of at most $\frac{1}{2}r$ from x. Such a subcomplex is contained in a face of a single simplex of Y, and is thus contractible.

Suppose that these is a vertex $v \in L$ with $\mathrm{d}_X(x, v) > \frac{1}{2}r$; we may take v attaining a maximal such distance. Let z be a point on a geodesic $[x, v] \subseteq X$ with $\mathrm{d}_X(z, v) = \frac{1}{2}r$ and a point $v' \in X'$ with $\mathrm{d}_X(z, v') \leq 1$. We will show that if $u \in L$ is a vertex with $\mathrm{d}_X(u, v) \leq r$, then $\mathrm{d}_X(u, v') \leq r$ also. This implies that if $(v, x_1, \dots, x_n)$ is a simplex Y, then so is $(v', x_1, \dots, x_n)$. Let $u \in L$ be a vertex with $\mathrm{d}_X(u, v) \leq r$. By the four-point inequality,

$$\mathrm{d}_X(u, v') + \mathrm{d}_X(x, v) \leq \max\{\mathrm{d}_X(u, v) + \mathrm{d}_X(x, v'), \mathrm{d}_X(u, x) + \mathrm{d}_X(v, v')\} + 2\delta.$$

In either case, one can use the defining inequalities from the previous paragraph to show that $\mathrm{d}_X(u, v') \leq \frac{1}{2}r + 2\delta + 1$. Now since $r \geq 4\delta + 2$, this implies $\mathrm{d}_X(u, v') \leq r$, proving the claim. Thus the subcomplex L' obtained by replacing every simplex with v as a vertex with v' is well-defined, and homotopic to L via the obvious affine maps in $(v, v', x_1, \dots, x_n)$. □

It is straightforward to apply the above to a hyperbolic group via its Cayley graph.

THEOREM 7.3. *Let G be a hyperbolic group. There is a simplicial complex P with:*

(1) P is finite dimensional, contractible, and locally finite;

(2) G acts on P simplicially, cocompactly, with finite cell stabilisers;

(3) the action is free and transitive on the vertex set of P

In particular, if G is torsion-free, then P/G is a finite classifying space for G.

PROOF. Take X to be a Cayley graph of G with respect to some finite generating set S. Then X is δ-hyperbolic for some δ, and $X = N_1(G)$, where G is viewed as the vertex set of X. Pick $r = 4\delta + 2$, so that by Proposition 7.2, $P = P_r(G)$ is contractible.

Given $g \in G$ there are at most $2|S|^r$ elements $h \in G$ with $\mathrm{d}_X(g, h) \leq r$, so that P is necessarily finite dimensional and locally finite. The vertex set of P is exactly G, which G acts on by (left) translation. This action is free and transitive, and extends to P by linearly interpolating across simplices. The action on P is necessarily simplicial, as it preserves adjacency in X. Finally, if $\sigma = (x_1, \dots, x_n)$ is a simplex of P, then G permutes $x_1, \dots, x_n$ faithfully. Hence the stabiliser of σ has order at most $n! = |S_n|$. □

Note that the above theorem tells us that P is what is known as a *classifying space for proper actions* for any hyperbolic group G, often denoted $\underline{E}G$, which is like an EG but with finite cell stabilisers. It just so happens that when G is torsion free, there are no finite subgroups, so that any $\underline{E}G$ as actually an EG.

CHAPTER 4

Groups acting on trees

1. Free constructions

There are some natural constructions in the category of groups that allow one to build larger groups out of smaller ones. Of course, among these are things like direct products (which is the categorical product) and, more generally, group extensions. We are interested here in the more free constructions of this variety. The most basic of these operations is the free product.

Definition 1.1 (Free product). Let G and H be groups. The *free product* $G*H$ of G and H is the coproduct of G and H. That is, for any group K and any homomorphisms $\varphi\colon G \to K, \psi\colon H \to K$, there is a unique homomorphism $f\colon G * H \to K$ such that the following commute:

$$\begin{tikzcd} G \arrow[r] \arrow[dr, "\varphi"'] & G * H \arrow[d, dashed, "f"] & H \arrow[l] \arrow[dl, "\psi"] \\ & K & \end{tikzcd}$$

Similarly to free groups, it is clean to define a free product in terms of a universal property, but it is often useful to have a model to work with (and to show that a coproduct actually exists!). We say a word in $G \sqcup H$ is *reduced* if it contains no consecutive pairs of the form gg' with $g, g' \in G$ or hh' with $h, h' \in H$. That is, it strictly alternates between letters in G and letters in H. There is an obvious reduction relation, and we can verify that the group of equivalence classes of reduced words in $G \sqcup H$ (with the operation of concatenation of representatives) is in fact the free product $G * H$.

Exercise 1.2. Show that if $G = \langle S \,|\, Q\rangle$ and $H = \langle T \,|\, R\rangle$ are presentations, then $G * H$ has the presentation $\langle S \cup T \,|\, Q \cup R\rangle$.

It can be useful to consider these free constructions in the context of topological spaces. The Seifert–van Kampen theorem tells us that the fundamental group of the wedge of two locally contractible spaces is a free product of the fundamental groups of the two spaces. Of course, the wedge and the free product are both categorical coproducts (in the category of pointed topological spaces and groups respectively), and the π_1 map is functorial. More generally, this theorem tells us that if we glue two spaces together along some open path-connected subspace, the fundamental group is a *pushout* of the corresponding fundamental groups. This brings us to the more general form of the free product.

Definition 1.3 (Amalgamated free product). Let G, H, and K be groups, and suppose that $\varphi\colon K \to G, \psi\colon K \to H$ are injective homomorphisms. The *free product of*

G and H amalgamated over K is the pushout of φ and ψ. Namely, it is the group $G *_K H$ such that for any group homomorphisms $G \to L$ and $H \to L$, there is a unique homomorphism $G *_K H \to L$ such that the following diagram commutes:

$$\begin{tikzcd}
K \arrow[r, "\varphi"] \arrow[d, "\psi"'] & G \arrow[d] \arrow[ddr, bend left] & \\
H \arrow[r] \arrow[drr, bend right] & G *_K H \arrow[dr, dashed] & \\
 & & L
\end{tikzcd}$$

In a slight abuse of notation, we usually suppress mention of the homomorphisms φ and ψ entirely, and treat K as a common subgroup of both G and H. Again, amalgamated free products have an obvious presentation.

Exercise 1.4. In the notation of the above definition, show that $G *_K H$ has the presentation of $G * H$ with the added relations that $\varphi(k) = \phi(k)$ for each $k \in K$.

Show that if G and H are finitely presented and K is finitely generated, then $G *_K H$ is finitely presented.

Like in a free product, elements in amalgamated free products can also be written in a unique minimal way as reduced words in G and H, though their description is a little more involved. We call such expressions *normal forms* for elements. That such a normal form exists once one fixes a transversal of the amalgamating subgroup in each of the factors is a consequence of the following theorem. Note that in a free product, the amalgamating subgroup is trivial, so that the transversals comprise the entirety of each factors, and hence every reduced word is already a normal form.

Theorem 1.5. *Let G and H be groups and let $K \leqslant G, H$ be a common subgroup. Let $a_1, \dots, a_n \in G *_K H$ that alternate between images of either G or H. If $a_1 \dots a_n = 1$, then either*

(1) $n = 1$ and $a_1 = 1$;
(2) $n > 1$ and there is $i = 1, \dots, n$ such that a_i is in the image of K.

The other main free construction of importance to infinite groups is known as the *HNN extension*, named after its inventors Graham Higman, Bernard Neumann, and Hannah Neumann. It is a little harder to describe this construction with a universal property (it is a homotopy colimit), so we give the traditional presentation-based definition.

Definition 1.6 (HNN extension). Let G be a group, $H \leqslant G$ be a subgroup, and $\varphi \colon H \to G$ an injective homomorphism. The *HNN* extension $G*_\varphi$ is the group with the presentation

$$\langle G, t \,|\, tht^{-1} = \varphi(h) \text{ for all } h \in H \rangle.$$

We call the subgroups H and $\varphi(H)$ the *associated subgroups* of the HNN extension, G the *base* of the extension, and t is called the *stable letter.*

Of course, an HNN extension is in general not an actual group extension. Again in analogy with spaces, HNN extensions correspond to fundamental groups of *partial mapping tori* — spaces one obtains by gluing pieces of another space to itself, along a cylinder say. This goes some way to explaining why there is no simple universal property for this construction, since there is no interval object in the category of groups. There are also normal forms for elements in an HNN extension, similarly to amalgamated free products. This is a consequence of the following.

THEOREM 1.7 (Britton's lemma). *Let $G*_\varphi$ be an HNN extension of G with associated subgroups H and $\varphi(H)$, with stable letter t. Let $g_0, \dots, g_n \in G$ and $\varepsilon_1, \dots, \varepsilon_n = \pm 1$. Suppose that $g_0 t^{\varepsilon_1} g_1 t^{\varepsilon_2} \dots t^{\varepsilon_n} g_n = 1$. Then either*

(1) $n = 0$ and $g_0 = 1$;

(2) $n > 0$ and for some $1 \leq i \leq n-1$, we have $\varepsilon_i = -\varepsilon_{i+1}$ and either $g_i \in H$, if $\varepsilon_i = 1$, or otherwise $g_i \in \varphi(H)$, if $\varepsilon_i = -1$.

EXAMPLE 1.8. Let G be a group, and $\varphi\colon G \to G$ an automorphism. Then the HNN extension $G*_\varphi$ with associated subgroups G and $\varphi(G) \cong G$ is exactly the semi-direct product $G \rtimes_\varphi \mathbb{Z}$, where $\mathbb{Z}$ is the infinite cyclic subgroup generated by the stable letter.

EXAMPLE 1.9. The Baumslag–Solitar group $\mathrm{BS}(m,n)$ is an HNN extension $\mathbb{Z}*_\varphi$, where φ is an isomorphism of the subgroups $m\mathbb{Z}$ and $n\mathbb{Z}$ of $\mathbb{Z}$.

It is a straightforward consequence of Britton's lemma that the natural inclusion of the base group into an HNN extension is an embedding. As such, HNN extensions are a particularly useful tool for building embeddings of groups. For instance, they play a significant role in the proof of Higman's theorem, which states that a finitely generated group is recursively presented if and only if it embeds into a finitely presented group. We give an simpler example of such an application.

THEOREM 1.10. *Every countable group embeds in a group generated by two elements.*

PROOF. Let $C = \{c_n \mid n \geq 0\}$ be a countable group, and let $F = C * \langle a, b\rangle$ be the free product of C with the free group on a and b. For simplicity, we assume that $c_0 = 1$ is the identity in C. Now the set $\{b^i a b^{-i} \mid i \geq 0\}$ freely generates an infinite rank free group H in $\langle a, b\rangle$, and similarly $\{c_i a^i b a^{-i} \mid i \geq 0\}$ also freely generates an infinite rank free group K in C. Take $G = F*_\varphi$ to be an HNN extension of F, where $\varphi\colon H \to K$ is such that $\varphi(b^i a b^{-i}) = c_i a^i b a^{-i}$ for each $i \geq 0$. Of course, C is embedded in F, which is in turn embedded in G, so C embeds in G. Moreover, G has the presentation

$$G = \langle F, t \mid tat^{-1} = b, tb^i a b^{-i} t^{-1} = c_i a^i b a^{-i}, i \geq 1\rangle,$$

from which it can be seen that a and t form a generating set. ☐

The above allows one to equip any countable group with a proper metric, as a subspace of a 2-generated group it embeds in with respect to its word metric. It should not be immediately obvious that this is possible, but that it is allows one to study the coarse geometry of countable groups.

EXERCISE 1.11. Use Britton's lemma to show that every finite order element of an HNN extension $G*_\varphi$ is conjugate into G.

2. Decompositions of groups

We will see that the free constructions above play a crucial role in understanding the algebraic structure of infinite groups. First, we will need a definition.

Definition 2.1 (Ends of spaces and groups). Let X be a topological space, and $K_1 \subseteq K_2 \subseteq \ldots$ a sequence of nested compact subsets, the union of whose interiors covers X. An *end* of X is a nested sequence $U_1 \supseteq U_2 \supseteq \ldots$, with each U_i a connected component of $X - K_i$. If G is group with finite generating set S, then $e(G)$ is the number of ends of the space $\Gamma(G, S)$.

The ends of space are straightforwardly seen to be independent of the choice of exhaustion (K_i). As a consequence, the space of ends is a quasi-isometry invariant among proper metric spaces. It follows also that the ends of a group do not depend on the choice of generating set. For a hyperbolic group G, the ends are exactly the connected components of the boundary ∂G. The following may be reminiscent of a similar fact we saw for the cardinalities of convergence groups.

Exercise 2.2. Show that a finitely generated group always has either $0, 1, 2$, or infinitely many ends.

Ends of groups were introduced independently by Freudenthal and Hopf. It is obvious that the finitely generated groups with zero ends are exactly the finite groups, as the Cayley graph of every finitely generated infinite group contains is unbounded. Freudenthal and Hopf both also obtained the following characterisation of the two-ended groups.

Theorem 2.3. *Let G be a group with two ends. Then G contains an infinite cyclic group of finite index.*

The above admits a great variety of different proofs. While most are not especially difficult or long, they are also not particularly easy or short, so we omit the proof here. In the 60s, Stallings obtained the following striking result, which can be interpreted as one of the first major theorems in geometric group theory.

Theorem 2.4 (Stallings). *A finitely generated group G has more than one end if and only if G is an amalgamated free product $G = H *_K H'$ where $K \neq H, H'$ is a finite subgroup, or an HNN extension $G = H*_\varphi$ with finite associated subgroups.*

3. Bass–Serre theory

Amalgamated free products and HNN extensions can be viewed as the most basic examples of a much larger and unified theory of combination constructions, which is intimately linked to the theory of group actions on trees. This theory is commonly called *Bass–Serre theory*; it arose initially in Serre's study of the group $\mathrm{PSL}_2(\mathbb{Q}_p)$ — a p-adic linear group whose Bruhat–Tits building is a tree — and was expanded on by Bass.

Since its inception, Bass–Serre theory has become an increasingly essential tool in geometric group theory and low-dimensional topology. A frequent use is to decompose given infinite groups into simpler pieces glued together in certain ways, which are often easier to understand individually.

We will need a little bit of terminology to be able to proceed. Recall that for us, a graph Γ is a set of *vertices* $V\Gamma$ and *edges* $E\Gamma$, together with initial and terminal vertex maps $\iota, \tau \colon E\Gamma \to V\Gamma$, and en edge inversion map $\bar{\cdot} \colon E\Gamma \to E\Gamma$, which has the property that $\bar{\bar{e}} = e$, $\iota(\bar{e}) = \tau(e)$ and $\tau(\bar{e}) = \iota(e)$.

Definition 3.1 (Graph of groups). A *graph of groups* is a tuple $\mathcal{G} = (\Gamma, G_-, \varphi_-)$ where Γ is a connected graph, G_v and G_e are groups for each $v \in V\Gamma$ and $e \in E\Gamma$, and $\varphi_e \colon G_e \to G_v$ is an injective homomorphism for each $e \in E\Gamma$. We require that for each $e \in E\Gamma$, we have an equality of groups $G_e = G_{\bar{e}}$. We call the groups G_v the *vertex groups*, G_e the *edge groups*, and φ_e the *edge morphisms*.

Similarly to how one defines the fundamental group of a space as the set of loops on a given basepoint up to homotopy, we can define a fundamental group of a graph of groups as the set of loops in the underlying graph, with the extra data that paths can pick up group elements from vertex spaces and that 'homotopy' will respect edge morphisms. One should think of this construction as gluing together the vertex groups along the edge groups in a way prescribed by the edge morphisms.

Definition 3.2 (Fundamental group of a graph of groups). Let $\mathcal{G} = (\Gamma, G_-, \varphi_-)$ be a graph of groups, and pick a basepoint $v_0 \in \Gamma$. Let $F(\mathcal{G})$ be the quotient of the free product $F(E\Gamma) * (*_{v \in V\Gamma} G_v)$, subject to the relations $e\bar{e} = \bar{e}e = 1$ and $e\varphi_e(g)\bar{e} = \varphi_{\bar{e}}(g)$ for all $e \in E\Gamma$ and $g \in G_e$.

The *fundamental group* $\pi_1(\mathcal{G}, v_0)$ of $\mathcal{G}$ based at v_0 is the subgroup of $F(\mathcal{G})$ consisting of (images of) words of the form $g_0 e_1 g_1 \dots e_n g_n$ where $e_1 \dots e_n$ forms a loop based at v_0, and $g_i \in G_{v_i}$ for each $i = 0, \dots, n$, with $v_i = \tau(e_i)$ when $i > 0$.

Of course, the isomorphism type of the fundamental group of a graph of groups is independent of the choice of basepoint in the underlying graph. Indeed, similarly to fundamental groups of spaces, changing the basepoint amounts to conjugating in the auxiliary group $F(\mathcal{G})$. We will thus often suppress mention of the basepoint and simply write $\pi_1\mathcal{G}$. We can now realise the constructions of the previous sectuin as exactly the fundamental groups of one-edge graphs of groups.

Example 3.3 (Amalgamated free product). Let Γ be the graph consisting of a single edge e with distinct endpoints u and v. Let G_u, G_v, and G_e be groups with injective homomorphisms $\varphi_e \colon G_e \to G_u$ and $\varphi_{\bar{e}} \colon G_e \to G_v$, and consider the graph of groups $\mathcal{G} = (\Gamma, G_-, \varphi_-)$. Picking u as a basepoint, the only loops in Γ are powers of $e\bar{e}$. Then

$$F(\mathcal{G}) \cong \langle G_u, G_v, e \,|\, e\varphi_e(g)e^{-1} = \varphi_{\bar{e}}(g), g \in G_e \rangle,$$

so that $\pi_1\mathcal{G}$ is exactly the subgroup generated by G_u and eG_ve^{-1}, which is easily seen to be isomorphic to $G_u *_{G_e} G_v$.

Example 3.4 (HNN extension). Let Γ be the graph consisting of a single edge e with $v = \iota(e) = \tau(e)$. That is, Γ is a single edge loop on a vertex v. Let G_v be a group, $G_e \leqslant G_v$ a subgroup and $\varphi_e \colon G_e \to G_v$ an injective homomorphism. Let $\mathcal{G} = (\Gamma, G_-, \varphi_-)$ be the corresponding graph of groups, considering $\varphi_{\bar{e}} \colon G_e \to G_v$ as the inclusion map of G_e as a subgroup. The loops in Γ are exactly the powers of e, so in this case $\pi_1\mathcal{G} = F(\mathcal{G})$. Hence we have the presentation $\pi_1\mathcal{G} = \langle G_v, e \,|\, ege^{-1} = \varphi_e(g), g \in G_e \rangle$, from which we can exactly see that $\pi_1\mathcal{G} \cong G_v *_{\varphi_e}$.

Remark 3.5. The fundamental group of any graph of groups can be realised as an iterated sequence of amalgamated free products and HNN extensions, by collapsing a single edge or a loop at a time. Thus statements about fundamental groups of graphs of groups can be proven by induction, with these single-edge cases as the base cases.

Fundamental groups of graphs of groups also admit normal forms, similarly to amalgamated products and HNN extensions. As in those settings, an immediate consequence is that the vertex groups canonically embed into the fundamental group.

THEOREM 3.6. *Let $\mathcal{G} = (\Gamma, G_-, \varphi_-)$ be a graph of groups, and suppose that $e_1 \dots e_n$ is a loop in Γ based at $v_0 \in V\Gamma$. If $g_0 e_1 g_1 \dots e_n g_n = 1$ in $\pi_1(\mathcal{G}, v_0)$, where $g_i \in G_{v_i}$ and $v_i = \tau(e_i)$, then either*

(1) $n = 0$ and $g_0 = 1$; or

(2) there is $i = 1, \dots, n-1$ such that $e_i = \overline{e_{i+1}}$ and $g_i \in \varphi_{e_i}(G_{e_i})$.

EXERCISE 3.7. Show that if $\mathcal{G}$ is a graph of groups, any finite subgroup of $\pi_1\mathcal{G}$ is conjugate into a vertex or edge group of $\mathcal{G}$.

The universal cover of a graph, in the traditional sense, is a tree, and the action of the fundamental group of the graph on this tree via deck transformations recovers the original graph. Analogously, one can build a sort of equivariant universal covering tree for a graph of groups, which admits an action of its fundamental group that will recover the original graph of groups as a quotient. The action of a fundamental group of a graph via deck transformations on its universal cover is free — it has no fixed points. By contrast, the action of a fundamental group of a graph of groups on its universal covering tree will in general have many point stabilisers, which will exactly recover the vertex and edge groups of the original graph of groups.

DEFINITION 3.8 (Bass–Serre tree). Let $\mathcal{G} = (\Gamma, G_-, \varphi_-)$ be a graph of groups. We construct a graph $T_\mathcal{G}$ with a $\pi_1\mathcal{G}$-action, called the *Bass–Serre tree* of $\mathcal{G}$.

The vertices of $T_\mathcal{G}$ will be the cosets of vertex groups of $\mathcal{G}$ in $\pi_1\mathcal{G}$, and the edges of $T_\mathcal{G}$ will likewise be cosets of edge groups of $\mathcal{G}$ in $\pi_1\mathcal{G}$. We must define the incidence relations and inversion map. Given an edge $gG_e \in ET_\mathcal{G}$, where $e \in E\Gamma$, we define $\overline{gG_e} = gG_{\bar{e}}$, $\iota(gG_e) = gG_{\iota(e)}$, and $\tau(gG_e) = gG_{\tau(e)}$. The $\pi_1\mathcal{G}$ action on $T_\mathcal{G}$ is given by the permutation action on cosets that comprise the vertices and edges of $T_\mathcal{G}$.

It is straightforward to check that this action preserves the graph structure, and that the stabilisers of vertices (respectively, edges) are exactly the conjugates in $\pi_1\mathcal{G}$ of the vertex (respectively, edge) groups of the graph of groups $\mathcal{G}$. By definition, there is one orbit of vertices for each vertex group of $\mathcal{G}$ and one orbit of edges for each edge group of $\mathcal{G}$. Moreover, an edge of $T_\mathcal{G}$ is incident to a vertex of $T_\mathcal{G}$ if and only if the corresponding cosets of the edge group and vertex group are defined on an edge incident to a vertex in Γ. It follows that, as a graph, $T_\mathcal{G}/\pi_1\mathcal{G}$ is isomorphic to Γ.

We will not prove it here, but the Bass–Serre tree of a graph of groups is indeed a tree in the sense that it is a connected graph with no non-trivial cycles. This construction has an obvious converse.

DEFINITION 3.9 (Quotient graph of groups). Let G be a group and T a tree. Suppose that G acts on T without edge inversions — that is, there is no $e \in ET$ and $g \in G$ such that $ge = \bar{e}$ – so that $\Gamma = T/G$ is a well-defined graph.

Let $s\colon \Gamma \to T$ be a section of the quotient map $T \to \Gamma$. For each $v \in V\Gamma$, let G_v be the stabiliser in G of $s(v) \in VT$, and likewise for each $e \in E\Gamma$, define G_e to be the stabiliser of $s(e) \in E\Gamma$. Moreover, the edge morphisms φ_e are defined as the inclusion maps $G_e \hookrightarrow G_v$ of edge stabilisers into their adjacent vertex stabilisers (after possibility conjugating, if one of the endpoints of an edge is outside the image of the section s). The resulting graph of groups $\mathcal{G} = (\Gamma, G_-, \varphi_-)$ is called the *quotient graph of groups* of T.

Note that acting without edge inversions is not a serious restriction, since we can always subdivide an edge if there are edge inversions. The fundamental structure theorem of Bass–Serre theory tells us that the these above constructions cohere; if G acts on a tree T, then T is in fact the Bass–Serre tree of for the quotient graph of groups of T.

THEOREM 3.10. *Let G be a group acting on a tree T without edge inversions, and let $\mathcal{G}$ be the quotient graph of groups. Then G is isomorphic to $\pi_1\mathcal{G}$ and there is a G-equivariant isomorphism between T and the Bass–Serre tree $T_{\mathcal{G}}$ of $\mathcal{G}$.*

EXAMPLE 3.11. The modular group $G = \mathrm{SL}_2(\mathbb{Z})$ admits an action on the hyperbolic plane $\mathbb{H}^2$ by Möbius transformations. Its fundamental domain is a triangle with one vertex at infinity, and the group acts by translations and inversions permuting a tiling of $\mathbb{H}^2$ by copies of this triangle. Take the arc C between the midpoint of the finite side of this triangle and one of the adjacent vertices. The graph formed by the translates $G \cdot C$ is a graph T with two orbits of vertices and one orbit of edges. As T consists of the union of all finite edges of the tiling and each triangle has only one such edge, T is in fact a tree. The vertex stabilisers are $\mathbb{Z}/6\mathbb{Z}$ and $\mathbb{Z}/4\mathbb{Z}$, and the edge stabiliser is $\mathbb{Z}/2\mathbb{Z}$. It follows that G is isomorphic to the amalgamated free product $\mathbb{Z}/6\mathbb{Z} *_{\mathbb{Z}/2\mathbb{Z}} \mathbb{Z}/4\mathbb{Z}$.

EXERCISE 3.12. Show that the fundamental group of a graph of finite groups has a finite index free subgroup. In particular, $\mathrm{SL}(2, \mathbb{Z})$ has a finite index free subgroup.

The correspondence given by Bass–Serre theory can also be used to give some sleek proofs of appealing statements.

THEOREM 3.13. *If G acts freely on a tree, then G is a free group.*

PROOF. Let G act on a tree T freely, and write $\Gamma = T/G$. The action is without edge inversion, as otherwise it would fix a midpoint of an edge. Then by the structure theorem $G \cong \pi_1\mathcal{G}$, where $\mathcal{G}$ is the quotient graph of groups. Since the action of G was free, the vertex and edge groups of $\mathcal{G}$ are trivial. It follows that $\pi_1\mathcal{G} \cong \pi_1\Gamma$ is a free group. □

The following states that a subgroup of a free product is again a free product, whose factors are a free group and conjugates of subgroups of the original free factors. It is rather a pain to prove this without Bass–Serre theory.

THEOREM 3.14 (Kurosh subgroup theorem). *Let G and H be groups, and let K be a subgroup of the free product $G * H$. Then there are collections of subgroups $\{G_i \mid i \in I\}$ and $\{H_j \mid j \in J\}$, conjugate into G and H respectively, and a subset X of $G * H$ such that*

$$K = \Big(\mathop{\ast}_{i \in I} G_i \Big) * \Big(\mathop{\ast}_{j \in J} H_j \Big) * F(X).$$

PROOF. Let T be the Bass–Serre tree of the free product $G * H$, viewed as the fundamental group of the single-edge graph of groups with trivial edge group and vertex groups G and H. Now $G * H$ acts on T without edge inversion, so K does as well. Let T_K be a minimal K-invariant subtree of T under this action. Then we see that $K = \pi_1\mathcal{K}$, where $\mathcal{K}$ is the quotient graph of groups of T_K. The edge groups are trivial, since the edge stabilisers in T were trivial. It follows that K is a free product of the vertex groups of $\mathcal{K}$ together with $\pi_1\Gamma$, where $\Gamma = T_K/K$. The vertex groups of $\mathcal{K}$ are all contained in vertex stabilisers of the action of $G * H$ on T, so they are conjugate into G or H as required. $\square$

4. Accessibility and further decompositions

The dictionary between graphs of groups and actions on trees provided by Bass–Serre theory gives us a powerful language to speak about groups with. For instance, the rather clunky statement of Stallings' theorem on ends of groups may be restated thus:

THEOREM 4.1. *If G is a finitely generated group with more than one end, then G acts non-trivially on a tree with finite edge stabilisers.*

If the vertex stabilisers of the action given by the above theorem are again many-ended, then one can apply the theorem once more to obtain a more refined decomposition of the original group. In theory, this process may never terminate; to say it does means that there is a finite graph of groups with finite edge groups and zero- or one-ended vertex groups.

DEFINITION 4.2 (Accessibility). Let G be a finitely generated group. We say that G is *accessible* if it acts on a tree T with finite edge stabilisers, vertex stabilisers with at most one end, and T/G a finite graph.

One would hope that this is true for every finitely generated group, but there are many examples, the first of which was constructed by Dunwoody, of finitely generated groups that are not accessible. Nevertheless, most countable groups one might care about are accessible.

THEOREM 4.3 (Dunwoody, '85). *Every finitely presented group is accessible.*

In particular, we saw that hyperbolic groups are finitely presented, hence accessible. Note that the statement of Dunwoody is actually a little more general: being finitely presented is in some sense a homotopical condition, and the statement holds for groups satisfying a (strictly weaker) homological analogue of finite presentability.

To prove the above, one constructs finite cut sets on the universal cover of a Cayley complex that are in some sense minimal. Dual to this collection of cut sets is a tree that is acted on by the group in the appropriate way.

Knowing that a group is accessible reduces many questions one could ask about that group to its finitely many one-ended 'factors'. For hyperbolic groups, the vertex groups in a graph of groups with finite edge groups are in fact quasiconvex, and so are also hyperbolic. Thus many things one could ask about the class of hyperbolic groups reduces to questions about one-ended hyperbolic groups (that is, hyperbolic groups with connected boundaries). For these, the next least complicated possible decompositions after splittings over finite groups are also understood: those where the edge groups are two-ended. Again, one can detect the splitting directly from connectedness properties of the boundary; recall that a *local cut point* of a topological space is one that disconnects at least one of its neighbourhoods.

THEOREM 4.4 (Bowditch, '97). *A one-ended hyperbolic group G acts on a tree with two-ended edge stabilisers if and only if ∂G contains a local cut point.*

The proof involves constructing a tree out of the local cut point structure of ∂G, which admits a canonical G-action. This resulting splitting is usually known as the *JSJ decomposition* of G, in rough analogy with an identically named decomposition in the theory of 3-manifolds.