Hyperbolic Groups
Lecture Notes

James Howie
Heriot-Watt University,
Edinburgh EH14 4AS
Scotland
J.Howie@ma.hw.ac.uk

Introduction

This is a revised and slightly expanded set of notes based on a course of 4 lectures given at the postgraduate summer school 'Groups and Applications' at the University of the Aegean in July 1999. I am very grateful to the University of the Aegean - in particular to Vasileios Metaftsis - for the opportunity to give the lectures and to publish the notes, and for their warm hospitality during the summer school.

The subject matter is hyperbolic groups - one of the main objects of study in geometric group theory. Geometric group theory began in the 1980's with work of Cannon, Gromov and others, applying geometric techniques to prove algebraic properties for large classes of groups. In this, the subject follows on from its ancestor, 'Combinatorial Group Theory', which has roots going back to the 19th century (Fricke, Klein, Poincaré). It adds yet another layer of geometric insight through the idea of treating groups as metric spaces, which can be a very powerful tool.

In a short lecture course I could not hope to do justice to this large and important subject. Instead, I aimed to give a gentle introduction which would give some idea of the flavour. I have tried to prepare these notes in the same spirit. One difficulty one faces when approaching this subject is the fact that there are several (equivalent) definitions. To give them all would be time-consuming, to prove equivalence more so. I have settled for giving just two definitions, each motivated by the corresponding geometric properties of the hyperbolic plane, and ignoring the question of equivalence.

The first lecture deals in general with groups seen as metric spaces, introduces the idea of quasi-isometry, and illustrates the ideas using the study of growth of groups. The second lecture gives the 'thin-triangles' definition of hyperbolic group, and uses it to give a simple proof that hyperbolic groups are finitely presented. (This is a consequence of a more far-reaching result of Rips, to which we return later.) The third lecture introduced Dehn diagrams and isoperimetric inequalities, gives the 'linear isoperimetric inequality' definition of hyperbolic groups, and indicates how to use this to obtain solutions of the word and conjugacy problems for hyperbolic groups. The final lecture was designed to give a glimpse of two slightly more advanced aspects of the subject, namely the Rips complex and the boundary of a hyperbolic group. In practice, I ran out of time and settled for discussing only the Rips complex. However, I have included a section on the hyperbolic boundary in these notes for completeness.

I hope that these notes will encourage readers to learn more about the subject. The principal references in this area are the original texts of Gromov [7, 8, 9], but several authors have worked on producing more accessible versions. I found [1, 2, 4] useful sources.

Lecture 1: Groups as metric spaces

Geometric group theory is the study of algebraic objects (groups) by regarding them as geometric objects (metric spaces). This idea seems unusual at first, but in fact is very powerful, and enables us to prove many theorems about groups that satisfy given geometric conditions.

How can we make a group G into a metric space? Choose a set S of generators for G. Then every element of G can be expressed as a *word* in the generators: $g = x_1^{\epsilon_1} x_2^{\epsilon_2} \ldots x_n^{\epsilon_n}$, where $x_1, x_2, \ldots, x_n \in S$ and $\epsilon_1, \epsilon_2, \ldots, \epsilon_n = \pm 1$. The natural number n is called the *length* of this word. If $g, h \in G$ then we define $d_S(g,h)$ to be the length of the *shortest* word representing $g^{-1}h$.

Lemma 1 *d_S is a metric on the set G.*

Proof. By definition, $d_S(g,h) \in \mathbb{N}$. In particular, $d_S(g,h) \geq 0$. Moreover, $d_S(g,h) = 0$ if and only if $g^{-1}h$ is represented by the empty word (of length 0). But the empty word represents the identity element of G, so $d_S(g,h) = 0 \Leftrightarrow g = h$.

If $x_1^{\epsilon_1} x_2^{\epsilon_2} \ldots x_n^{\epsilon_n}$ is a word of minimum length representing $g^{-1}h$, then $h^{-1}g = x_n^{-\epsilon_n} \ldots x_2^{-\epsilon_2} x_1^{-\epsilon_1}$, so $d_S(h,g) \leq d_S(g,h)$. Similarly, $d_S(g,h) \leq d_S(h,g)$, so $d_S(h,g) = d_S(g,h)$.

Let $x_1^{\epsilon_1} x_2^{\epsilon_2} \ldots x_n^{\epsilon_n}$ and $y_1^{\delta_1} y_2^{\delta_2} \ldots y_m^{\delta_m}$ be words of minimum length representing $g^{-1}h$ and $h^{-1}k$ respectively. Then

$$x_1^{\epsilon_1} x_2^{\epsilon_2} \ldots x_n^{\epsilon_n} y_1^{\delta_1} y_2^{\delta_2} \ldots y_m^{\delta_m}$$

is a word (not necessarily of minimum length) representing $g^{-1}h \cdot h^{-1}k = g^{-1}k$. Hence

$$d_S(g,k) \leq d_S(g,h) + d_S(h,k)$$

(the triangle inequality). ∎

Remarks

1. The metric d_S on G is called the *word metric* on G with respect to S. It takes values in $\mathbb{N}$. This distinguishes it from the metrics associated to more standard geometric objects (euclidean or hyperbolic space, surfaces, manifolds), which take values in $\mathbb{R}_+$. However, if the units of measurement are very small, or equivalently if we look at G 'from a great distance', then we cannot distinguish between the discrete-valued metric d_S and some continuous-valued approximation to it. This can all be made precise and used to compare G with more familiar and continuous metric spaces such as euclidean or hyperbolic space.

2. The metric d_S is related to the Cayley graph $\Gamma(G,S)$ in a natural way: we can identify G with the set of vertices of $\Gamma(G,S)$, and two vertices g,h ($g \neq h$) are adjacent in Γ if and only if $g^{-1}h \in S$ or $h^{-1}g \in S$, in other words if and only if $d_S(g,h) = 1$. More generally, if g,h are joined by a path of length n in $\Gamma(G,S)$, then we can express $g^{-1}h$ as a word of length n in S, so $d_S(g,h) \leq n$. The converse is also true: if $g^{-1}h$ can be expressed as a word of length n in S, then g,h can be joined by a path of length n in $\Gamma(G,S)$. Hence $d_S(g,h)$ is precisely the length of a shortest path (a *geodesic*) in $\Gamma(G,S)$ from g to h.

3. The metric d_S depends in an essential way on the choice of generating set S. For example, if we take $S = G$ then d_S is just the *discrete metric*: $d_S(g,h) = 1$ whenever $g \neq h$. This is not an interesting metric, and can not be expected to give interesting algebraic information about G. To avoid this kind of problem, we restrict attention to *finite* generating sets S. In particular, all groups from now on will be *finitely generated*.

4. Even with the restriction to finite generating sets, the metric depends on the choice of S. In particular, for any $g,h \in G$, we can choose a generating set S that contains $g^{-1}h$, so that $d_S(g,h) \leq 1$. However, despite such obvious problems, the dependence can be shown to be limited in a very real sense, so that if we look at G 'from a distance' then the effects of changing generating set become less apparent. In other words, there are many properties of the metric space (G, d_S) that are independent of the choice of S. These properties are the objects of study in geometric group theory.

Quasi-isometry

An *isometry* from one metric space (X,d) to another metric space (X',d') is a map $f : X \to X'$ such that

$$d'(f(x), f(y)) = d(x,y) \quad \forall x,y \in X.$$

It follows that f is continuous and injective. If f is also surjective, then $f^{-1} : X' \to X$ is also an isometry, and in this case we say that the metric spaces (X,d) and (X',d') are *isometric*. This is an equivalence relation between metric spaces. Isometric metric spaces are regarded as being 'the same', just as isomorphic groups or rings, or homeomorphic topological spaces, are 'the same'.

Quasi-isometry is a weaker equivalence relation between metric spaces that meets the requirements of geometric group theory by neglecting fine detail and concentrating on the large picture 'seen from a distance' as mentioned in the remarks above. It is defined in an analogous way. Let λ, k be positive real numbers. A map $f : X \to X'$

is a (λ, k)-*quasi-isometry* if

$$\frac{1}{\lambda}d(x,y) - k \leq d'(f(x), f(y)) \leq \lambda d(x,y) + k \ \ \forall x, y \in X.$$

Now f is not in general continuous or injective. If f is almost surjective, in the sense that every point of X' is a bounded distance from the image of f, then there is a (λ', k')-quasi-isometry $f' : X' \to X$ for some λ', k', that is almost an inverse to f. In this case we say that the metric spaces (X, d) and (X', d') are *quasi-isometric.*

Examples

1. $(\mathbb{Z}, d)$ and $(\mathbb{R}, d)$ are quasi-isometric, where d is the usual metric: $d(x, y) = |x - y|$. The natural embedding $\mathbb{Z} \to \mathbb{R}$ is an isometry, so a $(1, 0)$-quasi-isometry. It is not surjective, but each point of $\mathbb{R}$ is at most $\frac{1}{2}$ away from $\mathbb{Z}$. We can define a $(1, \frac{1}{2})$-quasi-isometry $f : \mathbb{R} \to \mathbb{Z}$ by $f(x) = x$ 'rounded to the nearest integer'.

2. We can generalise the above example. Let G be a group with a finite generating set S, and let $\Gamma = \Gamma(G, S)$ be the corresponding Cayley graph. We can regard Γ as a topological space in the usual way, and indeed we can make it into a metric space by identifying each edge with a unit interval $[0, 1] \subset \mathbb{R}$ and defining $d(x, y)$ to be the length of the shortest path joining x to y. This coincides with the path-length metric d_S when x and y are vertices. Since every point of Γ is in the $\frac{1}{2}$-neighbourhood of some vertex,we see that (G, d_S) and $(\Gamma(G, S), d)$ are quasi-isometric for this choice of d.

3. Every bounded metric space is quasi-isometric to a point.

4. $\mathbb{Z} \times \mathbb{Z}$ is quasi-isometric to the euclidean plane $\mathbb{E}^2$.

5. If S and T are finite generating sets for a group G, then (G, d_S) and (G, d_T) are quasi-isometric. Indeed, let λ be the maximum length of any element of S expressed as a word in T or *vice versa.* Then the identity map $G \to G$ is a $(\lambda, 0)$-quasi-isometry form (G, d_S) to (G, d_T) and *vice versa.* Hence, when we are discussing quasi-isometry in the context of finitely generated groups, we can omit mention of the particular generating set, and make statements like 'G is quasi-isometric to H' without ambiguity.

Growth

Suppose that G is a finitely generated group, and that S is a finite generating set for G. We define the *growth function* $\gamma = \gamma_S : \mathbb{N} \to \mathbb{N}$ for G with respect to S by

$$\gamma(n) = |\{\ g \in G \mid d_S(g, 1) \leq n\ \}|.$$

In other words, $\gamma(n)$ is the number of points contained in a ball of radius n in (G, d_S). The growth function of a finitely generated group clearly depends on the choice of generating set, but only in a limited way. Suppose that S and T are two finite generating sets for a group G. Let k be an integer such that every element of S can be expressed as a word of length k or less in T. Then for each integer n, the n-neighbourhood of 1 in G (with respect to the metric d_S) is contained in the kn-neighbourhood of 1 (with respect to d_T). Hence

$$\gamma_S(n) \leq \gamma_T(kn) \quad \forall n \in \mathbb{N}.$$

Similarly, there is an integer k' such that

$$\gamma_T(n) \leq \gamma_S(k'n) \quad \forall n \in \mathbb{N}.$$

Thus the asymptotic behaviour of $\gamma_S(n)$ as $n \to \infty$ is independent of S. This asymptotic behaviour is what is known as the *growth* of G

Similar arguments show the following.

Lemma 2 *Let G be a finitely generated group, H a subgroup of finite index, and γ, δ the growth functions of G, H respectively, with respect to suitable choices of finite generating set. Then there exists a constant $C > 0$ such that*

$$\gamma(n) \leq \delta(Cn), \quad \delta(n) \leq \gamma(Cn) \quad \forall n \in \mathbb{N}.$$

Thus the asymptotic behaviour of the growth function is the same for a subgroup of finite index.

More generally, if G and H are quasi-isometric finitely generated groups, then the asymptotic growth rates of G and H (with respect to any choice of finite generating sets) are the same. In other words, the asymptotic growth rate is a *quasi-isometry invariant*.

Examples

1. If G contains an infinite cyclic subgroup of finite index, then G has linear growth. It is enough to consider the growth function of $G = \mathbb{Z}$ with respect to the standard generating set $S = \{1\}$. But clearly $\gamma_S(n) = 2n + 1$, a linear function of n.

2. If G contains a free abelian group of rank r as a subgroup of finite index, then the growth of G is polynomial of degree r. Again, it is enough to consider $G = \mathbb{Z}^r$, with S a basis. A simple calculation shows that $\gamma_S(n)$ is a polynomial of degree r in n.

3. If G contains a free subgroup of rank greater than or equal to 2, then G has exponential growth. To see this, first note that, when S is a basis for a free group F of rank r, then the number of elements of F of length m in S is $2r(2r-1)^{m-1}$ (for all $m \geq 1$). Summing over $m = 1, \ldots, n$, we see that γ_S is exponential in n. Since every finitely generated group is a homomorphic image of a free group of finite rank, no group can grow faster than a free group. Thus no group grows faster than exponentially. Conversely, if G contains a free group F of rank $r \geq 2$, then we can choose a finite generating set T for G such that T contains a subset S that is a basis for F. Then $\gamma_T(n) \geq \gamma_S(n)$ grows exponentially.

The definitive result on growth of groups is the following, due to M Gromov [6]. (See also [10] for a survey on groups of polynomial growth, and [3] for an alternative proof of Gromov's Theorem.)

Theorem 1 *Let G be a finitely generated group. Then G has polynomial growth if and only if G has a subgroup of finite index that is nilpotent.*

There exist groups whose growth functions are *intermediate* (faster than any polynomial, but slower than any exponential). The first examples of these were due to R I Grigorchuk [5].

On the other hand, it is known that any group with growth bounded above by a polynomial function actually has polynomial growth. The degree of the polynomial can be computed from the lower central series of the nilpotent group.

Here is the simplest nonabelian example.

Example The *Heisenberg group* is the group H of 3×3 matrices with integer entries of the form

$$\begin{pmatrix} 1 & x & y \\ 0 & 1 & z \\ 0 & 0 & 1 \end{pmatrix}$$

It is nilpotent of class 2 with centre $Z(H) = [H, H]$ infinite cyclic. Its growth is polynomial of degree 4. To see this, we choose a system of three generators $\{a, b, c\}$, where

$$a = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad b = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}, \quad c = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Here $c = [a, b]$ is central in H: $[a, c] = [b, c] = 1$. Now it is not difficult to show that every element of H has a unique 'normal form' expression if the form $a^\alpha b^\beta c^\gamma$, $\alpha, \beta, \gamma \in \mathbb{Z}$. A naïve deduction from this would be that H has the same growth rate as $\mathbb{Z}^3$, which is cubic. However, for any $m, n \in \mathbb{Z}$ we have $[a^m, b^n] = c^{mn}$, so that the length of c^n as a shortest word in the generators grows asymptotically as

$4\sqrt{n}$, rather than n. (On the other hand, a^n and b^n are shortest words representing these elements for all n.) Hence the number of words of length n or less grows approximately like $n^4/4$.

Lecture 2: Thin triangles and hyperbolic groups

Hyperbolic groups are so-called because they share many of the geometric properties of hyperbolic spaces. There are a number of possible ways to define hyperbolic groups, which turn out to be equivalent. I will discuss only two of them. In order to motivate these, let us first look at some of the properties of the hyperbolic plane $\mathbb{H}^2$, where these differ substantially from the euclidean plane $\mathbb{E}^2$.

Thin triangles in the hyperbolic plane

The *incircle* of a triangle Δ (in $\mathbb{E}^2$ or $\mathbb{H}^2$) is the circle contained in Δ which is tangent to all three sides of Δ. The *inradius* of Δ is the radius of the incircle. For example, in $\mathbb{E}^2$ if Δ is an equilateral triangle whose sides have length ℓ, then the inradius of Δ is $\frac{\ell}{2\sqrt{3}}$. As $\ell \to \infty$, the inradius also tends to ∞.

The situation in $\mathbb{H}^2$ is quite different, however. Suppose $\Delta = \Delta_1$ is a triangle in $\mathbb{H}^2$ with vertices x, y, z, and let c be the *incentre* of Δ (that is, the centre of the incircle). Now for each $t \in \mathbb{R}_+$, let x_t be the point on the half-line from c through x, y, z respectively, such that $d(c, x_t) = t \cdot d(c, x)$. Define y_t, z_t in a similar way, and let Δ_t be the triangle whose vertices are x_t, y_t, z_t. The inradius of Δ_t is an increasing function of t. However, this time it does not tend to ∞ as $t \to \infty$. The limiting situation is an *ideal triangle* Δ_∞, whose vertices all lie on the boundary of $\mathbb{H}^2$. Now all ideal triangles in $\mathbb{H}^2$ are congruent, and have area π. This is clearly an upper bound for the area of the incircle of an ideal triangle, so 1 is an upper bound for the inradius of Δ_∞, and hence also for that of Δ.

In the hyperbolic plane $\mathbb{H}^2$, *all triangles are thin*, in the sense that there is a bound $\delta \in \mathbb{R}$ (actually, $\delta = 1$) such that the inradius of every triangle is less than or equal to δ. A consequence of this is that each edge of a hyperbolic triangle is contained in the 2δ-neighbourhood of the union of the other two edges: if x is a point on one edge of Δ, then there is a point y on one of the other edges of Δ such that $d(x, y) \leq 2\delta$.

Geodesic and hyperbolic metric spaces

In order to generalise the concept of thin triangles to other metric spaces, and hence to groups, we need to develop a more general notion of 'triangle'. A *geodesic segment* of length ℓ in a metric space (X, d) (from x to y) is the image of an isometric embedding $i : [0, \ell] \to X$ with $i(0) = x$ and $i(\ell) = y$. In other words, we have $d(i(a), i(b)) = b - a$ for all $0 \leq a \leq b \leq \ell$. A (*geodesic*) *triangle* Δ in X (with vertices x, y, z) is the union of three geodesic segments, from x to y, y to z and z to x respectively.

Note that in a metric space (X, d) there may not in general exist a geodesic segment from x to y (for example, if X is a discrete metric space). If geodesic segments exist, they need not be unique. For example, in $\mathbb{R}^2$ with the ℓ_1-metric

$$d_1((a_1, a_2), (b_1, b_2)) = |a_1 - b_1| + |a_2 - b_2|,$$

there are infinitely many geodesic segments from $(0, 0)$ to $(1, 1)$. A *geodesic metric space* is one in which there exist geodesic segments between all pairs of points. (There is no requirement for these geodesic segments to be unique.)

A geodesic metric space X is *hyperbolic* if all triangles are thin, in the following sense: there is a (global) constant δ such that each edge of each triangle Δ in X is contained in the δ-neighbourhood of the union of the other two sides of Δ.

Examples

1. Every bounded geodesic metric space is hyperbolic. If $d(x, y) \leq B$ for all x, y, then automatically any side of a triangle is contained in the B-neighbourhood of the union of the other two sides.

2. Every tree is a hyperbolic metric space. It is clearly geodesic, since any two points are connected by a shortest path. Moreover, any side of a triangle is contained in the union of the other two sides.

3. The hyperbolic plane $\mathbb{H}^2$ is a hyperbolic metric space, by the thin triangles property for $\mathbb{H}^2$ described above.

4. More generally, hyperbolic n-space $\mathbb{H}^n$ is a hyperbolic metric space by the thin triangles property for $\mathbb{H}^2$ (since any geodesic triangle is contained in a plane.

5. Euclidean space $\mathbb{E}^n$ is not a hyperbolic metric space for $n \geq 2$, since $\mathbb{E}^2$ does not satisfy the thin triangles property.

Lemma 3 *Let (X, d) and (X', d') be geodesic metric spaces that are quasi-isometric to one another. If (X, d) is hyperbolic, then so is (X', d') (and conversely).*

The proof of this is not difficult, but is quite technical, so I will omit it. Details can be found, for example, in [4, p. 88]. The key point is that hyperbolicity for metric spaces is an invariant of quasi-isometry type, which is important because the metric space defined by a finitely generated group is only well-defined up to quasi-isometry.

Hyperbolic groups

If G is a group, with (finite) generating set S, then (G, d_S) is not a geodesic metric space, since d_S takes values in $\mathbb{N}$. However, the geometric realization of the Cayley graph $K = |\Gamma(G, S)|$ *is* geodesic, with respect to the natural metric (which is quasi-isometric to G). If K is hyperbolic as a metric space, then G is called a *hyperbolic group*.

The first thing to note is that this property is independent of choice of generating set, since being hyperbolic is a quasi-isometry invariant. Also, subgroups of finite index in hyperbolic groups are hyperbolic. Conversely, groups containing subgroups of finite index that are hyperbolic are themselves hyperbolic.

Examples

1. Every finite group is hyperbolic, because its Cayley graphs are all bounded.
2. Every free group is hyperbolic, because it has Cayley graphs that are trees. Moreover, if G has a free subgroup of finite index, then G is quasi-isometric to a free group, and hence hyperbolic.
3. The fundamental group of a surface of genus $g \geq 2$ is quasi-isometric to $\mathbb{H}^2$, and hence is hyperbolic.
4. $\mathbb{Z} \times \mathbb{Z}$ is quasi-isometric to $\mathbb{E}^2$, and hence is not hyperbolic.

The fact that $\mathbb{Z} \times \mathbb{Z}$ is not hyperbolic shows that not every finitely generated group is hyperbolic. In fact, a stronger statement than this is true: there are $2^{\aleph_0}$ isomorphism classes of finitely generated groups, but only $\aleph_0$ of these are hyperbolic.

How do we know this? Not by examining an uncountable set of groups individually, but by a simple cardinality argument. The set of all *finite* group presentations is countable, by the standard countability argument. The following is a simple version of a theorem due to E Rips (see Théorème 2.3 on page 60 of [2]).

Theorem 2 *Every hyperbolic group is finitely presented.*

Proof. Let G be a hyperbolic group, and let $d = d_S$ be the metric on G determined by some *fixed* finite generating set S. For each $n \in \mathbb{N}$ we define

$$X_n = \{\ g \in G \mid \delta(g, 1) \leq n\ \}$$

and

$$R_n = \{\ xyz \mid x, y, z \in X_n,\ xyz = 1 \text{ in } G\ \} \cup \{\ xx^{-1} \mid x \in X_n\ \} \subset F(X_n).$$

Then

$$X_1 \subset X_2 \subset \dots$$

and

$$R_1 \subset R_2 \subset \dots$$

so if we define $G_n = \langle\, X_n \mid R_n \,\rangle$ then we obtain a sequence of group homomorphisms

$$G_1 \to G_2 \to \dots \to G_\infty = G.$$

Note first that these homomorphisms are surjective. If $g \in X_{k+1} \setminus X_k$, with $k \geq 1$, then there exist elements $u, v \in X_k$ with $uvg = 1$ in G. Since $u, v, g \in X_{k+1}$ it follows that $uvg \in R_{k+1}$, so $uvg = 1$ in G_{k+1}. Hence the image of $G_k \to G_{k+1}$ contains the generating set X_{k+1} and so it is surjective.

Next we show that $G_N \to G_{N+1}$ is injective (and hence an isomorphism) for all sufficiently large N. It follows that $G \cong G_N$ for all large N, or equivalently that $G = \langle\, X_N \mid R_N \,\rangle$, so is finitely presented, as claimed.

We fix $N \gg 2\delta$. Suppose that $xyz \in R_{N+1}$. In other words, $x, y, z \in X_{N+1}$ with $xyz = 1$ in G. We have to show that this relation can be deduced from those in R_N. Unfortunately, the elements x, y, z do not in general belong to X_N. To make sense of this, we first choose, for each $x \in X_{N+1} \setminus X_N$, a *canonical* splitting $x = x_1x_2$ with each of $d(x_1, 1) > \delta$, $d(x_2, 1) > \delta$ and $d(x, 1) = d(x_1, 1) + d(x_2, 1)$. We then add the generator x and the relation $x_1x_2x^{-1}$ to the presentation for G_N to get an equivalent presentation. Having done this, we now show how to deduce $xyz = 1$ from the relations in X_N together with the canonical splitting relations.

Case 1 $x, y \in X_N$, $z \notin X_N$.

Let P be the point of the geodesic segment z corresponding to the canonical splitting z_1z_2. By the thin triangle property, this is within distance δ of some point Q on one of the other edges of the geodesic triangle Δ with vertices at $1, x, xy$. The geodesic PQ, together with the geodesic from Q to the vertex of Δ opposite the edge containing Q, divides the geodesic triangle into three smaller triangles, each of the edges of which has length N or less. It follows that the relation $xyz_1z_2 = 1$ can be deduced from three relations in R_N, as required.

Case 2 $y, z \notin X_N$.

By case 1 we can assume all (true) relations of the form $abc = 1$ with $a, b \in X_N$ and $c \in X_{N+1}$. Here we proceed exactly as in case 1, starting from the canonical splitting of z. The point Q may lie on an edge of length $N + 1$, in which case it corresponds to a (possibly non-canonical) splitting of x or y - say $x = x_1'x_2'$ (but still with x_1', x_2' of lengths less than $N + 1$). Hence the relation $x = x_1'x_2'$ is one we are allowed to assume. Finally, we divide Δ as before into three smaller triangles. This time it is possible that one of the three triangles has one side of length $N + 1$, but all other sides of the smaller triangles have length N or less. By case 1 we are done.

Similar arguments apply to the relations of the form xx^{-1}, $x \in X_{N+1} \setminus X_N$, completing the proof.

■

Lecture 3: Isoperimetric inequalities and decision problems

Consider a Jordan curve C in the euclidean plane $\mathbb{E}^2$. The Jordan curve theorem tells us that C bounds a compact domain D in E^2. The *isoperimetric inequality* compares the area of D to the length of C. To make sense of this, let us assume that C is nice (smooth, polygonal, piecewise smooth, ...) so that it has a well-defined (finite) length and D has a well-defined (finite) area. A classical result of calculus of variations says that, for C of fixed length ℓ, the area of D is maximised when C is a circle (of radius $r = \ell/2\pi$). This maximal area is $\pi r^2 = \ell^2/4\pi$. Hence the isoperimetric inequality for $\mathbb{E}^2$ is:

$$\text{Area}(D) \leq \frac{\ell^2}{4\pi}.$$

Note that the right hand side of this inequality is a *quadratic* function of ℓ.

We can look at the same thing in the hyperbolic plane $\mathbb{H}^2$. Again the maximal area occurs when C is a circle. In $\mathbb{H}^2$ the length of a circle of radius r is $2\pi\sinh(r)$, and its area is

$$\int_0^r 2\pi\sinh(t)dt = 2\pi(\cosh(r) - 1) \leq 2\pi\sinh(r)$$

so in this case we have an isoperimetric inequality

$$\text{Area}(D) \leq \ell$$

for domains D bounded by curves of length ℓ. The important difference here is that the right hand side of the inequality is a *linear* function of ℓ.

What is the relevance of this for groups?

Let $\mathcal{P} : \langle\ X \mid R\ \rangle$ be a finite presentation of a group G. If $w \in F(X)$ is a word that represents the identity element $1 \in G$, then it can be expressed, in $F(X)$, as a product of conjugates of elements of R and their inverses:

$$w = (u_1^{-1}r_1^{\epsilon_1}u_1)\ldots(u_n^{-1}r_n^{\epsilon_n}u_n),$$

where $u_i \in F(X)$, $r_i \in R$ and $\epsilon_i = \pm 1$. There will in general be infinitely many such expressions. The *least* value of n amongst all such expressions is called the *area* of w, $\text{Area}(w)$.

The notion of area for words representing $1 \in G$ has a geometric interpretation.

Definition A *van Kampen diagram* or *Dehn diagram* over the presentation $\mathcal{P}$ consists of the following data:

- A simply-connected, finite 2-dimensional complex M contained in the plane.
- An orientation of each 1-cell of M.
- A labelling function that labels each 1-cell of M by an element of X, such that the composite label of the boundary of each 2-cell (read from a suitable starting point in a suitable direction) is an element of R. (Here, an edge labelled $x \in X$ contributes x to the boundary label of the 2-cell if read in the direction of its orientation, and x^{-1} if read in the opposite direction.)

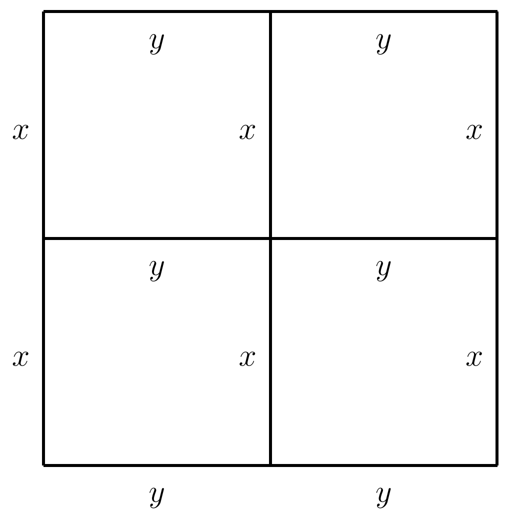
The complement of M in the plane is topologically a punctured disc. It also has a boundary that is a closed path in the 1-skeleton of M. The label of this path is called the *boundary label* of the diagram. Note that it is defined only up to cyclic permutation and inversion.

Lemma 4 *There exists a Dehn diagram with boundary label w if and only if $w = 1$ in G, in which case the minimum number of 2-cells in all Dehn diagrams with boundary label w is* $\mathrm{Area}(w)$*.*

Example $\mathcal{P} = \langle\ x,\ y,\ \mid [x, y]\ \rangle$, (where $[x, y]$ means $xyx^{-1}y^{-1}$),

$$w = [x^2, y^2] = (x[x, y]x^{-1})([x, y])(yx[x, y](yx)^{-1})(y[x, y]y^{-1}).$$

Then $\mathrm{Area}(w) = 4$.



The function $f : \mathbb{N} \to \mathbb{N}$ defined by

$$f(\ell) = \max\{\mathrm{Area}(w) \mid w \in F(X),\ w = 1 \text{ in } G,\ \ell(w) = \ell\}$$

is called the *Dehn function*, or *isoperimetric function* for the presentation $\mathcal{P}$. A given finitely presented group has (infinitely) many possible finite presentations, which can have very different Dehn functions. However, there are aspects of Dehn functions that are independent of the choice of presentation, and hence are invariants of the group G.

Lemma 5 *Let $\mathcal{P}$ and $\mathcal{Q}$ be two finite presentations for a group G, and let f, g be the corresponding Dehn functions. Then there exist constants $A, B, C, D \in \mathbb{N}$ such that*

$$f(n) \leq Ag(Bn + C) + D \quad \forall\ n \in \mathbb{N}.$$

In particular, if g is bounded above by a function that is linear (or quadratic, or polynomial, or exponential, ...) in n, then the same is true for f. These properties are thus invariants of the group G.

Definition A finitely presented group G has a *linear* (*quadratic, ...*) *isoperimetric inequality* if for some (and hence for any) finite presentation $\mathcal{P}$ with Dehn function f, there is a linear (quadratic, ...) function $\hat{f}$ such that $f(n) \leq \hat{f}(n)\ \forall\ n \in \mathbb{N}$. A finitely presented group G is *hyperbolic* if it has a linear isoperimetric inequality.

We have now given two distinct definitions for hyperbolic group. Implicit in this is an assertion that these two properties are equivalent: the thin triangles condition is equivalent to the linear isoperimetric inequality condition.

Examples

1. Every finite group is hyperbolic. If G is a finite group, then the Cayley table for G is a finite presentation for G. In other words, we take G to be the finite generating set, and the set of all true equations $xy = z$ in G for the set of defining relations. Given a word $w \in F(G)$ such that $w = 1$ in G, what is $\mathrm{Area}(w)$? If w has length 3, then it is a relation in G. If it has length greater than 3, then it has the form $w = xyu$, where $x, y \in G$ and u is a word. If $z \in G$ with $z = xy$, then $w = (xyz^{-1})(zu)$ with zu shorter than w. An inductive argument shows that $\mathrm{Area}(w) \leq \ell(w)$.

2. Every free group is hyperbolic. Indeed, if $F = \langle\ X \mid -\ \rangle$ is the standard presentation, then the empty word 1 is the only cyclically reduced word representing the identity in F, and $\mathrm{Area}(1) = 0$.

3. The fundamental group of a surface of genus $g \geq 2$ is hyperbolic.

   This is because of a Theorem of Dehn: let

$$G = \langle\ a_1, b_1, \ldots, a_g, b_g \mid [a_1, b_1] \cdots [a_g, b_g]\ \rangle$$

and let w be a word in the generators a_i, b_i such that $w = 1$ in G. Then there exist cyclic permutations $w' = uv$ of w and $r' = us$ of the relator $r = [a_1, b_1] \cdots [a_g, b_g]$ or its inverse r^{-1}, with a common initial segment u of length greater than half the length of r, that is $\ell(u) > 2g$. It follows that $\text{Area}(w) = \text{Area}(uv) \leq \text{Area}(s^{-1}v) + 1$, while $\ell(w) < \ell(s^{-1}v)$. An inductive argument then shows that $\text{Area}(w) < \ell(w)$.

4. $\mathbb{Z} \times \mathbb{Z}$ is not hyperbolic. Under the presentation $G = \langle\, x, y \mid [x, y] \,\rangle$, the word $w_n = [x^n, y^n]$ has area at most n^2. Indeed there is a Dehn diagram for w_n which is a square of side length n in $\mathbb{E}^2$, subdivided into n^2 squares of side length 1. On the other hand, the boundary of this Dehn diagram is a simple closed path in $\mathbb{E}^2$ bounding a square of area n^2, so n^2 is a lower bound for the area of w_n^2. Now $\ell(w_n) = 4n$, so this sequence of words shows that no linear isoperimetric inequality holds for $\mathbb{Z} \times \mathbb{Z}$.

The word problem

Let G be a group given by a (finite) presentation $\langle\, X \mid R \,\rangle$. (Much of what follows can also be done for certain types of infinite presentation, but let us keep things simple.) The *word problem* for G is that of deciding algorithmically whether or not a given word w in the generating set X represents $1 \in G$. A *solution* to the word problem is an algorithm that, when an arbitrary word w is input, will output **YES** or **NO** after a finite time, depending on whether or not $w = 1$ in G.

As with most of what we have been doing, this problem appears at first sight to depend on the choice of finite presentation for G. However, it can be shown to be independent of this choice. Indeed, given two finite presentations $\langle\, X \mid R \,\rangle$ and $\langle\, Y \mid S \,\rangle$ of isomorphic groups G and G', there are functions $X \to F(Y)$ and $Y \to F(X)$ that induce the isomorphisms. Given an algorithm to solve the word problem for $\langle\, X \mid R \,\rangle$ and a word w' in Y, we can apply the function $Y \to F(X)$ to rewrite w' as a word w in X, then apply our solution to decide whether or not $w = 1$ in G. Since this is true if and only if $w' = 1$ in G', we are done. (**NB** this solution assumes the *existence* of isomorphisms $G \leftrightarrow G'$. Although the practical implementation of the solution uses one of these isomorphisms, it does not assume that we are able to find it for ourselves. Indeed, the problem of determining whether or not too given presentations represent isomorphic groups is another insoluble decision problem, called the *isomorphism problem*).

At first sight the word problem seems trivial. We have a very good criterion for deciding whether or not $w = 1$ in G: namely, can we express w as a product of conjugates of elements of R (and their inverses) in $F(X)$? Why can we not use this criterion as an algorithm? In fact, this criterion supplies part of the answer, but the other part is missing, in general.

Since X and R are finite sets, the set $F(X)$ is countable, and hence so is the set of conjugates of elements of $R \cup R^{-1}$ by elements of $F(X)$. Hence so also is the set of all finite products of such conjugates. In principal, the standard proof that these sets are countable can be transformed into an algorithm to produce an infinite list that contains all products of conjugates of elements of $R \cup R^{-1}$ (possibly with repetitions). This can be combined with an algorithm to check each element of the list for equality with w. Once we recognise that w is in the list, we stop and output the answer **YES**.

However, what if $w \neq 1$ in G? Our algorithm will continue to list words forever, without ever recognising w. It will *never* give us the desired output **NO**! We may, after a long wait, *suspect* that $w \neq 1$, but we can never be completely sure. If our computer runs for 1000 years without producing an output, this could be for three reasons: $w \neq 1$; $w = 1$ but w appears much later than 1000 years in the computer-generated list; or possibly hardware error.

In fact, it is known that there are finitely presented groups with unsolvable word problem. In other words, there is no algorithm of the kind we are looking for. However, for hyperbolic groups there is indeed an algorithm. Because of the linear isoperimetric inequality, there is a bound on Area(w) for words w such that $w = 1$ in G, *depending only on* $\ell(w)$. Since R is finite, it can easily be shown that there are only finitely many Dehn diagrams of a given area and length of boundary, and hence these can be searched in finite time, comparing boundary labels with w. (**NB** it is not true that there are only finitely many Dehn diagrams of a given area. For example, if $r \in R$ then there is a Dehn diagram of area 2 for each of the infinitely many words $[r, u]$, $u \in F(X)$. The condition on boundary length is also important to make this search work.)

Lemma 6 *Hyperbolic groups have solvable word problem*

This result is not so strong as it might appear. In fact, our argument does not use the fact that hyperbolic groups have a *linear* isoperimetric inequality. All that is required is that the Dehn function be bounded above by some function that we can compute (a *recursive* function). This is a much weaker property.

Lemma 7 *Let G be a group given by a (finite) presentation with a recursive isoperimetric inequality. Then G has soluble word problem.*

The conjugacy problem

However, there are other types of decision problem where the stronger hyperbolic property can be successfully applied.

Let G be a group with finite presentation $\langle\ X \mid R\ \rangle$. A solution to the *conjugacy problem* for G is an algorithm which, when two words w, w' in X are input, will determine in finite time whether or not w, w' represent conjugate elements of G. As before, this can be shown to be independent of the choice of presentation.

Lemma 8 *If G is hyperbolic, then the conjugacy problem is soluble for G.*

Proof. The idea of the proof is roughly as follows. Suppose that w, w' are words that represent conjugate elements of G, and that the conjugating element $g \in G$ is represented by some *very long* word u. Then the relation $w'uw^{-1}u^{-1} = 1$ in G corresponds to a van Kampen diagram of area bounded above by an *a priori* given linear function f of $\ell(u)$. Let $x_0, \ldots, x_{\ell(u)}$ and $y_0, \ldots, y_{\ell(u)}$ denote the sequences of vertices along the boundary of the diagram corresponding to the two segments labelled u. (So that the paths from x_0 to $x_{\ell(u)}$ and from y_0 to $y_{\ell(u)}$ are both labelled u). Then, roughly speaking, $d(x_i, y_i)$ is bounded above by some constant multiple of the constant function $f'(\ell(u))$, where d denotes the path-length metric in the 1-skeleton of the Dehn diagram.

Hence there is a constant N such that, whenever $\ell(u) > N$, two of the pairs (x_i, y_i) and (x_j, y_j) of vertices are joined by paths p_i, p_j of equal lengths and with equal labels. Cutting the diagram open along p_i and p_j, and then regluing it by identifying p_i with p_j, we get a diagram corresponding to a new relation $w'vw^{-1}v^{-1} = 1$ with $\ell(v) < \ell(u)$. Hence we can restrict our search for conjugating elements to those of length less than N. Since there are only finitely many of these, it suffices to apply the solution to the word problem to a finite number of words. ■

The key part of this proof is the notion of the finite bound for the 'width' of the long strip that corresponds to the Dehn diagram for $w'uw^{-1}u^{-1} = 1$. The details of this part of the argument are tricky. The bound has to be calculated without specific reference to w, w', except that it depends on $m = \ell(w)$ and $n = \ell(w')$.

For fixed m, n, the area of the potential diagrams under consideration will be bounded by a linear function of $\ell(u)$. Since we are working with a given finite presentation for G, the lengths of relations are also bounded, and hence the number of edges in the diagram is bounded by a linear function of $\ell(u)$. If the geodesic paths connecting the x_i to the y_i are not of bounded length, then the sums of their lengths will grow faster than linearly with $\ell(u)$, leading to arbitrarily complicated intersections between these paths for very large u, and hence to a contradiction.

An alternative approach to proving this theorem can be found, for example, in [2, pages 52-56].

Lecture 4: Further aspects

The Rips Complex

We have already seen that hyperbolic groups are finitely presented, a result I attributed to Rips. It is in fact a special case of a much stronger theorem of Rips, which I will try to explain in this section. Topologically, the fact that a group G is finitely presented means that there is a finite simplicial complex Y with $G \cong \pi_1(Y)$. In fact, this condition is both necessary and sufficient for G to be finitely presented. Moreover, we have a lot of freedom in the choice of Y. For example, we can choose Y to be 2-dimensional, since the fundamental group of a simplicial complex is determined by its 2-skeleton (the union of all the simplices of dimensions ≤ 2).

Let Y be a finite simplicial complex with $G \cong \pi_1(Y)$, and let K be its universal cover. Then K is a simply connected complex, and there is a simplicial action of G on K such that $Y = K/G$ is compact. The G-action is free ($gx = x \Rightarrow g = 1$), and so in particular it is properly discontinuous. The existence of a simply connected simplicial complex K and a G-action on K with these properties is yet another necessary and sufficient condition for G to be finitely generated. (Here the weaker properly discontinuous condition on the action is sufficient: one can then construct another simply connected complex K' on which G acts freely, again with K'/G compact.)

The *Rips complex* is a particular choice of K for a hyperbolic group G that is not only simply connected but contractible. Thus the result we want to prove is the following.

Theorem 3 *(Rips) Let G be a hyperbolic group. Then there exists a simplicial complex K and an action of G on K such that*

1. *G acts properly discontinuously on K;*

2. *K/G is compact;*

3. *K is contractible.*

For the proof, we will fix a finite generating set S for G, and a positive real number δ such that G is δ-hyperbolic. We will actually construct a simplicial complex $\mathcal{P}_n(G)$ on which G acts, for every natural number n.

Define
$$\mathcal{P}_n(G) = \{Y \subset G \mid Y \neq \emptyset \ \& \ \mathrm{diam}_S(Y) \leq n\}.$$

Some explanation is required here. Firstly, $\mathrm{diam}_S(Y)$ denotes the diameter of the set Y with respect to the metric d_S:

$$\mathrm{diam}_S(Y) = \max\{d_S(x, y) \mid x, y \in Y\}.$$

Secondly, what we have defined here is a collection of subsets of G. They are all finite subsets, since each $Y \in \mathcal{P}_n(G)$ is contained in a ball of radius n, which has only finitely many elements. They are nonempty sets by definition. We are implicitly identifying a $(k+1)$-element set Y with a k-simplex whose vertex set is Y. Clearly

$$\emptyset \neq X \subset Y \in \mathcal{P}_n(G) \Rightarrow X \in \mathcal{P}_n(G),$$

so our collection of simplices $\mathcal{P}_n(G)$ is closed with respect to faces. It also follows from this that two simplices of $\mathcal{P}_n(G)$ are either disjoint or intersect in a common face of both. These are the defining properties for simplicial complexes, so $\mathcal{P}_n(G)$ is indeed a simplicial complex, as claimed.

There is a natural action of G on $\mathcal{P}_n(G)$: if $g \in G$ and $Y = \{x_0, \ldots, x_k\} \in \mathcal{P}_n(G)$, then

$$Yg = \{x_0 g, \ldots, x_k g\} \in \mathcal{P}_n(G).$$

Since G is acting by permuting simplices, the action is simplicial.

We next note that the first two properties in Rips' Theorem hold automatically for all natural numbers n.

Claim 1 *G acts properly discontinuously on $\mathcal{P}_n(G)$ for all n.*

To see this, note first that the set of vertices (or 0-simplices) of $\mathcal{P}_n(G)$ is just G (or, more correctly, the set of 1-element subsets of G), with G acting by right multiplication. If Y is a k-simplex and $g \in G$ with $Y \cap Yg \neq \emptyset$, then $x = yg$ for some $x, y \in Y$, so $g = y^{-1}x$ with $x, y \in Y$. There are only finitely many such g. Hence G acts properly discontinuously, as claimed.

Claim 2 *$\mathcal{P}_n(G)/G$ is compact for all n.*

It is enough to show that there are only finitely many orbits of simplices in $\mathcal{P}_n(G)$. If $Y = \{x_0, \ldots, x_k\} \in \mathcal{P}_n(G)$ then $Y' = Yx_0^{-1} \in \mathcal{P}_n(G)$ with $1 \in Y'$. But then Y' is contained in the n-neighbourhood of 1, which is finite. There are thus only finitely many possibilities for Y', so only finitely many orbits of simplices.

For the third property of $\mathcal{P}_n(G)$ the value of n is important. For example, $\mathcal{P}_0(G)$ is just the discrete set G regarded as a 0-dimensional simplicial complex, so it is not even connected (unless G is the trivial group). The point of Rips' Theorem is that the hyperbolic property of G ensures that $\mathcal{P}_n(G)$ is contractible *for large n*. To specify what we mean by large n, recall that we have fixed a metric $d = d_S$ on G, with respect to which G is δ-hyperbolic for some real number δ.

Proposition 1 *If $n > 4\delta + 2$, then $\mathcal{P}_n(G)$ is contractible.*

Proof. If not, then there is a nonzero homotopy group $\pi_t(\mathcal{P}_n(G)) \neq 0$, so a continuous map $f : S^t \to \mathcal{P}_n(G)$ that is not homotopic to a constant. The image of f is compact, so contained in a finite subcomplex K of $\mathcal{P}_n(G)$. It is therefore sufficient to show that every finite subcomplex K is contractible to a point 1 within $\mathcal{P}_n(G)$.

Let X denote the (finite) set of vertices of the (finite) subcomplex K of $\mathcal{P}_n(G)$. We argue by induction on

$$\nu = \sum_{x\in X} d(1,x).$$

Another useful parameter is

$$\mu = \max_{x\in X} d(1,x).$$

If $\mu \le n/2$ then for any $x, y \in X$ we have

$$d(x,y) \le d(x,1) + d(1,y) \le \frac{n}{2} + \frac{n}{2} = n,$$

so X is the vertex set of a simplex of $\mathcal{P}_n(G)$, which is therefore contractible (to 1), and K is a union of faces of this simplex. The result is therefore true in this case, which includes the initial case of the induction.

Suppose then that $\mu > n/2$, and choose $x_1 \in X$ with $d(1,x_1) = \mu$. There is a geodesic γ in the Cayley graph $\Gamma = \Gamma(G,S)$ from 1 to x_1, and we define $x_0 \in G$ to be the point on this geodesic with $d(x_0,x_1) = [n/2]$, the integer part of $n/2$. We now define a map $f : X \to G$ by

$$f(x) = \begin{cases} x & (x \ne x_1) \\ x_0 & (x = x_1) \end{cases}$$

Now $f(X)$ is a finite set whose ν parameter is less than that of X, so we can apply the inductive hypothesis to any subcomplex of $\mathcal{P}_n(G)$ with vertex set $f(X)$. To complete the inductive argument, it is sufficient to show that f extends to a simplicial map $f : K \to \mathcal{P}_n(G)$ that is homotopic to the identity map on K. If Δ is a simplex of K that does not contain x_1, then f is the identity on Δ, so it is enough to consider the action of f on simplices that contain x_1. Let Δ be such a simplex, and D its vertex set. I claim that $D \cup \{x_0\}$ is the vertex set of a simplex Δ' of $\mathcal{P}_n(G)$. Then Δ and $f(\Delta)$ will be faces of Δ', so f will be simplicial and homotopic to the identity on Δ.

To prove the existence of Δ', we need only check that any two of its vertices are no more than n apart in the metric $d = d_S$. This is true by definition for two vertices of Δ, so we need only prove that $d(x_0,y) \le n$ for every vertex $y \in \Delta$: in other words, we must prove that

$$[d(x_1,y) \le n \ \&\ d(1,y) \le \mu = d(1,x_1)] \Rightarrow d(x_0,y) \le n.$$

Choose geodesics α from 1 to y and β from y to x_1 in Γ. Then α, β, γ are the sides of a geodesic triangle in Γ, and x_0 is a point on γ. By the thin triangle property there is a point z on $\alpha \cup \beta$ such that $d(x_0,z) < \delta$.

Case 1: $z \in \alpha$.

Then $d(1, z) = d(1, y) - d(y, z) \leq \mu - d(y, z)$ and so

$$\mu = d(1, x_1) \leq d(1, z) + d(z, x_0) + d(x_0, x_1) \leq \mu - d(y, z) + \delta + [n/2].$$

Thus $d(y, z) \leq \delta + [n/2]$ and so

$$d(y, x_0) \leq d(y, z) + d(z, x_0) \leq \delta + [n/2] + \delta < n.$$

Case 2: $z \in \beta$.

Then $[n/2] = d(x_0, x_1) \leq d(x_0, z) + d(z, x_1) \leq \delta + d(z, x_1)$, so $d(z, x_1) \geq [n/2] - \delta$. Moreover, $d(y, z) + d(z, x_1) = d(y, x_1) \leq n$, so $d(y, z) \leq n - [n/2] + \delta \leq [n/2] + 1 + \delta$. Finally,

$$d(y, x_0) \leq d(y, z) + d(z, x_0) \leq [n/2] + 1 + \delta + \delta \leq n.$$

■

Boundary of a hyperbolic group

The ordinary hyperbolic plane, when considered in the Poincaré disc model, has a natural boundary - the 'circle at infinity'. Similarly, hyperbolic space of dimension n has a natural boundary which is an $(n-1)$-sphere. There are analogous constructions for geodesic metric spaces in general, and for hyperbolic groups in particular.

The purpose of this section is to describe this construction and an application to the structure of hyperbolic groups. We will omit most of the technical details for the benefit of expositional clarity. We refer the reader to [2, 4] for more rigorous treatments.

Consider first the hyperbolic plane $\mathbb{H}^2$. In the Poincaré disc model, this is the interior of the unit disc in $\mathbb{R}^2$: $\{x \in \mathbb{R}^2, ||x|| < 1\}$. The boundary of the disc, $\partial\mathbb{H}^2 = S^1 = \{x \in \mathbb{R}^2, ||x|| = 1\}$ is a compact space. Indeed the union

$$\bar{\mathbb{H}}^2 = \mathbb{H}^2 \cup \partial\mathbb{H}^2$$

is a compact space in which $\mathbb{H}^2$ is an open dense subset, so is a natural compactification of $\mathbb{H}^2$. There are various ways of thinking about the boundary S^1, any of which can be used to produce the analogous constructions for hyperbolic metric spaces in general.

I will describe only one of these ideas. Fix a basepoint $x_0 \in \mathbb{H}^2$. (In the Poincaré disc model, one should think of x_0 as being the euclidean origin $(0, 0)$.) The set of geodesic rays starting at x_0 is naturally identified with $S^1 \subset \mathbb{E}^2$. Each such ray meets the boundary in a unique point, so we get an identification $\partial\mathbb{H}^2 \cong S^1$. The advantage of this approach is that it induces a metric on $\partial\mathbb{H}^2$, using the euclidean or spherical metric on S^1. The metric depends on the choice of base-point, but the underlying topology does not.

From now on, we consider a hyperbolic group G, with a fixed generating set S, and corresponding Cayley graph $\Gamma = \Gamma(G, S)$. For our basepoint x_0, we make the canonical choice of the identity element $1 \in G = V(\Gamma)$. We then consider geodesic rays from x_0 (that is, isometries $[0, \infty) \to \Gamma$ with $0 \mapsto x_0$).

For some purposes, it is useful to be able to vary the starting point of the geodesic, so that we relax the condition $0 \mapsto x_0$.

Two geodesics g, h are said to be *equivalent* if $\{d(g(t), h(t)), t \in [0, \infty)$ is bounded. Then $\partial G = \partial\Gamma$ is defined to be the set of equivalence classes of geodesic rays.

Examples

1. If G is finite, then $\partial G = \emptyset$.

2. If G is an infinite cyclic group, then $\partial G = \{-\infty, +\infty\}$.

3. If G is a Fuchsian group, then the Cayley graph Γ embeds quasi-isometrically in $\mathbb{H}^2$. Every geodesic ray in Γ can be approximated by a geodesic ray in $\mathbb{H}^2$, and *vice versa*, so that there is a natural identification between ∂G and $\partial H^2 \cong S^1$.

4. Similarly, if G is a Kleinian group, then $\partial G = \partial\mathbb{H}^3 = S^2$.

5. If G is a nonabelian free group, then Γ is a tree and ∂G is a Cantor set.

Remark In the above examples, we have made no explicit mention of the particular generating set of G chosen to determine ∂G. It is not difficult to see that the definition of ∂G does not depend on the choice of generating set. Indeed, ∂G is an invariant of quasi-isometry, so for example $\partial G = \partial H$ if H is a subgroup of finite index in G.

Also implicit in the above examples is the existence of a natural topology on ∂G. In fact, we can define a metric on ∂G. As with the hyperbolic plane $\mathbb{H}^2$, the metric on ∂G will depend on our choice of base-point, although the resulting topology does not. This is important for what follows. We use the canonical choice of the identity element of G to be the base-point. Given two elements $x, y \in G$, we define the *Gromov inner product*

$$\langle x|y\rangle = \frac{d(1,x) + d(1,y) - d(x,y)}{2}.$$

A good way to think of this is as follows. Recalling the thin triangles property of hyperbolic metric spaces, we see that two travellers moving from 1 to x and y

respectively along suitable geodesics will remain close together (less than 2δ apart) for a certain distance, before beginning to diverge rapidly. The inner product $\langle x|y\rangle$ measures approximately the length of time that the two travellers remain close together. In particular, $\langle x|y\rangle = 0$ if and only if the travellers diverge immediately, that is, iff there is a geodesic from x to y that passes through 1.

The inner product extends to the boundary by taking limits as x and/or y vary along geodesic rays. The inner product on G is enough to recover the original metric, but we have to be slightly more subtle to define a metric on ∂G.

Definition Choose $\epsilon > 0$. Define $\rho_\epsilon(a, b) = exp(-\epsilon\langle a|b\rangle)$ for $a, b \in \partial G$, and

$$d_\epsilon(a, b) = \inf \sum_{i=1}^{n} \rho_\epsilon(a_{i-1}, a_i),$$

the infimum being taken over all finite sequences $a_0, \ldots, a_n \in \partial G$ with $a_0 = a$ and $a_n = b$.

Then d_ϵ is a metric on ∂G for sufficiently small ϵ. In general ρ_ϵ is not a metric, because it does not satisfy the triangle inequality. It turns out that $\bar{G} = \Gamma \cup \partial G$ is a compact topological space, such that the subspace topology induced on ∂G is the same as that induced by the metric d_ϵ.

To apply the boundary of G to obtain results about its algebraic structure, we use the fact that elements of G act by isometries on the Cayley graph Γ. This action extends to $\bar{G}$ as an action by homeomorphisms. Not isometries, as we have not defined a metric on $\bar{G}$! In general, the restriction of the action to ∂G is also not by isometries, since the metric on ∂G depends on the base-point, which is not fixed by the isometries of Γ in question.

Lemma 9 *Any element of G acts on Γ with bounded orbits if and only if it has finite order. Any element of infinite order in G fixes a point of ∂G.*

Proof. Clearly, if $g \in G$ has finite order n, then its orbits have the form

$$\{h, gh, \ldots, g^{n-1}h\},$$

and so are bounded. Conversely, an orbit of g is bounded if and only if it is finite, which can happen if and only if g has finite order. If g has infinite order, then the sequence $1, g, g^2, \ldots$ can be approximated by a geodesic ray, and so tends to a limit $a \in \partial G$. ∎

We can apply this fact to the study of centralisers in G. Suppose x has infinite order, and $a = \lim_{n\mapsto\infty} x^n \in \partial G$ as in the proof of the Lemma. If y belongs to the centraliser of x in G, then $yx^n = x^n y$ for all n, so

$$d(yx^n, x^n) = d(1, x^{-1}yx) = d(1, y)$$

is constant. It follows that y maps the geodesic ray that approximates the sequence $1, x, x^2 \ldots$ to one a bounded distance from the original (in other words, another geodesic ray in the same equivalence class). Hence y fixes $a \in \partial G$. We can therefore study the centraliser of G by analysing the dynamics of the stabiliser of a acting on ∂G.

From the thin triangles property, it is easy to deduce that there is in fact a global bound on the distance between two equivalent geodesic rays for large t, provided that one is allowed to vary the starting point of one of the rays. If D is an upper bound for $d(g(t), h(t))$, and $T \gg t \gg D$, then the point $h(t)$ on the geodesic for $h(0)$ to $h(T)$ must be close to some point y on the geodesic from $g(0)$ to $h(T)$ (since it is not close to any point of the geodesic from $h(0)$ to $g(0)$). Similarly, y is close to some point on the geodesic from $g(0)$ to $g(T)$, ie to $g(t+u)$ for some u. Once one analyses this argument, one obtains the following global bound. If g, h are equivalent rays, then there are real numbers C, u such that

$$d(g(t), h(t-u)) \leq 16\delta \forall t > C.$$

Returning to the study of centralisers, one can show that only finitely many of elements of the centraliser of x give rise to a given value of u in the above inequality, from which it follows that $\langle x \rangle$ has finite index in its centraliser.

Theorem 4 *The centraliser of an element x of infinite order in a hyperbolic group G contains the cyclic group $\langle x \rangle$ as a subgroup of finite index.*

Corollary 1 *No hyperbolic group contains a free abelian subgroup of rank 2.*

Corollary 2 *The group $\mathbb{Z} \times \mathbb{Z}$ is not hyperbolic*

Corollary 3 *The only knot group which is hyperbolic is the infinite cyclic group (the group of the unknot).*

Proof. The group of any nontrivial knot contains a *peripheral* subgroup (that is, the fundamental group of the boundary of a regular neighbourhood of the knot) isomorphic to $\mathbb{Z} \times \mathbb{Z}$. ■

References

[1] J Alonso *et al*, *Notes on negatively curved groups*, in: Group theory from a geometrical viewpoint (ed. E Ghys, A Haefliger and A Verjovsky), World Scientific, 1991, pp 3-63.

[2] M Coornaert, T Delzant and A Papadopoulos, *Géometrie et théorie des groupes (Les groupes hyperboliques de Gromov)*, Lecture Notes in Mathematics **1441**, Springer-Verlag, 1990.

[3] L van den Dries and A J Wilkie, *Gromov's theorem on groups of polynomial growth and elementary logic*, J Algebra **89** (1984), 349-374.

[4] E Ghys and P de la Harpe (editors), *Sur les groupes hyperboliques d'apres Mikhael Gromov*, Progress in Mathematics **83**, Birkhäuser, 1990.

[5] R I Grigorchuk, Milnor's problem on the growth of groups, Dokl. Akad. Nauk. SSSR **271** (1983), 30-33; Soviet Math. Dokl. **28** (1983) 23-26.

[6] M Gromov, *Groups of polynomial growth and expanding maps*, Pub. Math. IHES **53** (1981), 53-73.

[7] M Gromov, *Infinite groups as geometric objects* Proceedings of the International Congress of Mathematicians (Warsaw, 1983), PWN, Warsaw, 1984, pp 385-392.

[8] M Gromov, *Hyperbolic groups*, in: Essays in group theory (ed. S M Gersten), MSRI Publications **8**, Springer-Verlag, 1987, pp 75-265.

[9] M Gromov, *Asymptotic Invariants of Infinite Groups*, (Geometric Group Theory, volume 2, edited by G A Niblo and M A Roller), London Mathematical Society Lecture Notes Series **182**, Cambridge University Press, 1983.

[10] J Tits, *Groupes á croissance polynomiale (d'après M. Gromov et al.)*, Seminaire Bourbaki 1980/81, Lecture Notes in Math **901**, Springer (1981), pp 176-188.