Introduction to geometric group theory

July 23, 2017

Please bring to my attention any mistake or typo by writing to chloeperin@gmail.com.

Idea of geometric group theory: to understand the algebraic properties of a group, it is useful to see it as the group of symmetries of **something** (preferably with a geometric structure whatever that means).

Where do groups come from? Apart from "classic objects" such as $\mathbb{Z}, \mathbb{R}$, most groups we know are given as groups of symmetries (we will see in the sequel another way of defining groups, which is completely algebraic - group presentations).

Example:

1. S_n group of permutations on n elements, A_n alternating group;
2. D_k group of symmetries of a regular k-gon;
3. D_∞ group of symmetries of the marked real line (a subgroup of $\text{Isom}(\mathbb{R})$);
4. $GL_n(\mathbb{R})$ group of invertible linear transformations of the vector space $\mathbb{R}^n$; $O(n)$ orthogonal group (=group of isometries);
5. Group of symmetries of the cube;
6. Groups of symmetries of tilings of the plane: square tiling, triangle groups

Given any mathematical object O, we can define its group of symmetries - they are bijective maps from O to itself preserving the structure on the object. Eg: permutations if the object is a set, linear maps if it's a vector space, isometries if it's a metric space, homeomorphisms if it's a topological space, etc. If we endow the set of symmetries of O with the product given by composition, we get a group $\text{Sym}(O)$.

In fact historically groups first appeared in Galois' work when he tried to understand solutions of polynomial equations through the group of symmetries of the roots.

Klein's Erlangen program in 1872 claimed that the way to understand geometry is through groups, i.e. that each geometry (euclidean, affine, projective, ...) can be understood by looking at its group of symmetries.

Geometric group theory, which arose in the late 1980's, can be thought of as implementing Erlangen program in reverse: to understand a group, one should understand what geometric objects it acts on (where the term "geometric" is taken in a loose sense here). Thus one asks two types of questions:

1. If I know a group G admits an action with properties xyz on a space of type T, what does it tell me about G?
2. Does the group G have an action with properties xyz on a space of type T?

Thus a typical question of geometric group theory: is $Aut(\mathbb{F}_n)$ linear? i.e. something given as a group of symmetries of a group, can it be thought of as a group of symmetries of a vector space? (It is not - for $n \geq 3$- Formanek and Procesi 1992)

1 Cayley graph and group presentations

Let us first think about the second question in much generality - i.e. without assuming anything on the group G.

Note that it is always possible to see a group as a group of permutations of a set - namely, the underlying set of the group.

Remark 1.1: *Let G be a group. Then G acts on itself by left translation, i.e. $g \cdot h = gh$ (check this is an action). The only element which acts like the identity permutation (fixes everyone) is the trivial element, hence the action is faithful. This gives an embedding $G \hookrightarrow \mathrm{Sym}(G)$ given by $g \mapsto \sigma_g$ where $\sigma_g(h) = gh$ (make sure you see why it's a morphism, and why injectivity corresponds to faithfulness of the action).*

We can do better than this: we will see that any group can be seen as a group of symmetries of an (oriented, labelled) graph - the Cayley graph.

1.1 Cayley graph

Definition 1.2: *Let G be a group, let S be a generating set for G such that $1 \notin S$. The Cayley graph of G with respect to S is an oriented graph whose vertices are the elements of G and where there is an edge (g, gs) for each $g \in G$ and $s \in S$. We can label the edge (g, gs) by s, thus making $X(G,S)$ an oriented labelled graph.*

Note that if $s^2 = 1$ then (g, gs) and (gs, g) are both edges in $X(G,S)$. We often take the convention to identify them to form one unoriented edge. We sometimes forget the orientation, or the labeling.

Example 1.3:
- $\mathbb{Z}$ with generating sets $\{1\}$, $\{2,3\}$;
- $\mathbb{Z}^2$ with standard generating set;
- $\mathbb{Z}/n\mathbb{Z}$;
- S_3 with respect to $(12), (123)$;
- D_n with respect to a rotation and a reflection. D_∞.

Lemma 1.4: *Cayley graphs are connected (as non oriented graphs) and regular.*

Proof. Connected because S generates. Regular: each vertex has an incoming edge labelled s and an outcoming edge labelled s (these two edges are identified if $s^2 = 1$). The valence is therefore $2|S| - \left|\{s \in S \mid s^2 = 1\}\right|$ □

Remark 1.5: *The group G acts on $X(G,S)$: if $h \in G$, we associate to it the symmetry of $X(G,S)$ which sends the vertex g to the vertex hg for each $g \in G$, and the edge (g, gs) to the edge (hg, hgs) for each $g \in G$ and $s \in S$. This action is faithful. Note that the action preserves orientation and labelling of the graph.*

Example 1.6: Action of $\mathbb{Z}$, $\mathbb{Z}^2$ on their various Cayley graphs.

Exercise 1.7: Draw the Cayley graph of $(\mathbb{Z}/2\mathbb{Z})^2$, and of $(\mathbb{Z}/2\mathbb{Z})^3$ relative to the standard generating sets, and understand the action of the group on the Cayley graph.

The geometric (or topological) realization of a graph is the quotient space formed by taking a copy of the interval $[0,1]$ for each edge and identifying common endpoints. It is a topological space. In fact, we even have a natural metric on this space induced by setting each edge to be isometric to the interval $[0,1]$ (easy to define the length of a path, then take distance between two points to be inf of length of paths joining them).

Remark 1.8: *The action of G on the graph $X(G,S)$ gives an action of G on the geometric realization of $X(G,S)$ by isometries. From now on blur the distinction between $X(G,S)$ and its geometric realization.*

Lemma 1.9: *If S does not contain elements of order 2, the action is free (i.e. no non trivial element fixes a point).*

Proof. Suppose $g \cdot x = x$. If x is a vertex, $x \in G$ and $g \cdot x = gx$, thus $gx = x$ implies $g = 1$.

If x is on an edge between vertices h and hs, then g must stabilize this edge - if it fixes the endpoints, as above this means $g = 1$. If it swaps them, we have $gh = hs$ and $ghs = h$ so $hss = h$ so $s^2 = 1$. $\square$

Example 1.10: Action of $\mathbb{Z}/2$ on its Cayley graph not free.

1.2 Words and paths

Definition 1.11: *Let S be a set. We introduce a symbol s^{-1} (the formal inverse) for each symbol $s \in S$, and set $S^{-1} = \{s^{-1} \mid s \in S\}$. A **word in** $S \cup S^{-1}$ is a finite sequence of elements of $S \cup S^{-1}$ (we will often abuse the terms and say a word in S). The empty word is usually denoted by e or by 1.*

Remark 1.12: *If S is a generating set for a group G, to a word w in $S \cup S^{-1}$ given by $s_1^{\epsilon_1} \dots s_k^{\epsilon_k}$ corresponds a unique path p_w in the Cayley graph starting at the vertex 1 and going along the edges labelled by the s_i, going with the orientation if $\epsilon_i = +1$ and against the orientation $\epsilon_i = -1$. Conversely any finite path in $X(G, S)$ starting at 1 gives us a word in $S \cup S^{-1}$.*

Example 1.13: In $\mathbb{Z}^2$, look at the words $aabbab^{-1}$, $bbaaa$, $aba^{-1}a^{-1}$.

The path p_w has endpoint at the vertex $g = s_1^{\epsilon_1} \cdot \dots \cdot s_k^{\epsilon_k}$ where $\cdot$ represents the product law of G and we take (obviously) s^{-1} to be the inverse of s in the group G.

Remark 1.14: *More often we write $g = s_1^{\epsilon_1} \dots s_k^{\epsilon_k}$ (omitting the product symbol), which can lead to some confusion. For example, if $S = \{a, b\}$ the **words** abb^{-1} and a are distinct, while the **elements of the group** abb^{-1} and a are the same. In the Cayley graph: the paths are distinct, but they have the same endpoints. Sometimes clarify by writing $abb^{-1} =_G a$.*

Definition 1.15: *We say that the word w **represents** the element g.*

Note that if we start the path at h instead of 1, you will reach the vertex hg.

Definition 1.16: *Let G be a group and let S be a generating set for G. The word length of an element g of G relative to S is*

$$l_S(g) = \min\{r \mid g = s_1^{\epsilon_1} \dots s_r^{\epsilon_r} \text{ for } s_i \in S \text{ and } \epsilon_i \in \{1, -1\}\}.$$

with the convention that $l_S(1) = 0$.

Note that this is exactly the length of the shortest path between 1 and g in the Cayley graph, that is (according to our choice of the metric in the geometric realization of a graph), the distance between 1 and g.

Definition 1.17: *Let G be a group and let S be a generating set for G. The word metric on G relative to S is*

$$d_S(g, h) = l_S(g^{-1}h)$$

In other words: by how many generators (or inverses of) do you need to multiply g to get h? Note that this is the distance between g and h in the Cayley graph.

Example 1.18: In $\mathbb{Z}^2$ with standard basis $\{a, b\}$, the path corresponding to $aba^{-1}b^{-1}$ is a loop - its endpoint is 1. This is because $aba^{-1}a^{-1} =_{\mathbb{Z}^2} 1$.

Definition 1.19: *If w is a word on S such that $w =_G 1$, we call w a relation between the generators S.*

Sometimes the equation $w =_G 1$ is called a relaTION, and w itself is called a relaTOR.

Note that in any group, ss^{-1} is a relation - these are not interesting.

Definition 1.20: *A word w is reduced if whenever $s_i = s_{i+1}$ we have $\epsilon_i = \epsilon_{i+1}$.*

In other words, a word is reduced iff the corresponding path in the Cayley graph has no backtrack - call it a reduced path. EXCEPT if have taken the convention of identifying edges corresponding to elements of S of order 2. In this case a path might have backtrack but correspond to a reduced word (eg ss for $s \in S$ with $s^2 = 1$).

1.3 Free groups

The idea of presentations is to define a group by giving a set of generators and some relations between these generators that should "imply" all the relations that exist. First, we consider the case where we want the generators to have as few relations as possible.

Definition 1.21: *Let G be a group, S a generating set. Say G is free on S if no non empty reduced word represents the trivial element.*

Remark 1.22: *Thus if G is free on S then $X(G,S)$ is a tree (it is a connected graph where there are no cycles without backtrack). If $X(G,S)$ is a tree and S contains no element of order 2 then G is free on S.*

Example 1.23: The subgroup F of $SL_2(\mathbb{Z})$ generated by the two matrices $\alpha = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}$ and $\beta = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}$ is free on $\{\alpha, \beta\}$. Note that for any $n \in \mathbb{Z}$ we have $\alpha^n = \begin{pmatrix} 1 & 2n \\ 0 & 1 \end{pmatrix}$ and $\beta^n = \begin{pmatrix} 1 & 0 \\ 2n & 1 \end{pmatrix}$

Consider the linear action of $\mathrm{SL}_2(\mathbb{Z})$ on $\mathbb{R}^2$. Let $U = \{(x,y) \mid |y| > |x|\}$, and $V = \{(x,y) \mid |y| < |x|\}$. Note that $\alpha^n(U) \subseteq V$ and $\beta^n(V) \subseteq U$ for any $n \in \mathbb{Z}$ with $n \neq 0$.

Let w be a non empty reduced word in α, β: assume first that w starts and ends by a power of α. Take $x \in U$: then $w \cdot x$ is in V, so it cannot be equal to x. Thus the element represented by w is not trivial.

Reduce to this case by conjugating w by an appropriate power of α. Show that the conjugate w' of w is non trivial, thus w itself is non-trivial.

This is a typical example of GGT proposition: look at action of the group on the space to deduce something algebraic. In fact, the technique we used here can be formalized to give the so-called ping-pong lemma (see later).

Remark 1.24: *In fact, a lot of linear groups (i.e. groups of matrices) contain free subgroup - Tits alternative (72) any finitely generated group of matrices either contains a free subgroup or is virtually solvable.*

So we know how to find free groups "in nature", if we have an appropriate action. But we can also build free groups abstractly.

Given a set S (of any cardinality), consider the set of words on S we define an equivalence relation by saying two words are equivalent if they can be obtained one from the other by a finite sequence of insertions or deletions of subsequences ss^{-1} or $s^{-1}s$.

Given two equivalence classes, choose representatives $w = u_1 \dots u_k$ and $w' = v_1 \dots v_l$: the product $[w][w']$ is defined to be the equivalence class $[ww']$ of the concatenation $ww' = u_1 \dots u_k v_1 \dots v_l$. It is possible to check that the equivalence class does not depend on the choice of the representatives w, w'.

Lemma 1.25: *Each class contains exactly one reduced word.*

It is not hard to see that the set of equivalence classes together with the product operation we defined forms a group that we call $F(S)$. Its neutral element is the class of the empty word $[\epsilon]$. The inverse of the class $[u_1 \dots u_k]$ is the class $[u_k^{-1} \dots u_1^{-1}]$.

Lemma 1.26: *The group $F(S)$ is free on S.*

Proof. We really mean "is free on $\{[s] \mid s \in S\}$"...Let $s_1^{\epsilon_1} \dots s_k^{\epsilon_k}$ be a non empty reduced word. We have $[s_1]^{\epsilon_1} \dots [s_k]^{\epsilon_k} = [s_1^{\epsilon_1} \dots s_k^{\epsilon_k}]$. Now we know that each class contains exactly one reduced element, hence this is not the trivial element (the only reduced word in the class corresponding to the trivial element is the empty word). This proves the result. □

In view of the Lemma above, we can think of an element in $F(S)$ as a reduced word on $S \cup S^{-1}$. The product of two reduced words is given by concatenation plus reduction.

Example 1.27: $S = \{a, b\}$, take $w = abb^{-1}a^{-1}$ and $w' = ab$ and reduce ww' to ab.

The following lemma is key. It says that it is very easy to define group morphisms with source $F(S)$.

Lemma 1.28: *Given any group G and a map $h : S \to G$ given by $s \mapsto g_s$ of G, there is a unique group morphism $H : F(S) \to G$ extending h.*

Idea: send the reduced word $s_1^{\epsilon_1} \dots s_k^{\epsilon_k}$ (which is an element in $F(S)$) on the product $g_{s_1}^{\epsilon_1} \dots g_{s_k}^{\epsilon_k}$ in G.

Proof. Formal approach - if $w = s_1^{\epsilon_1} \dots s_k^{\epsilon_k}$ send $[w]$ to $g_{s_1}^{\epsilon_1} \dots g_{s_k}^{\epsilon_k}$ - need to check this is well defined, i.e. that if we choose another representative we get the same image. Once we know this it is easy to show that this is a morphism.

$$H([w][w']) = H([ww']) = g_{s_1}^{\epsilon_1} \dots g_{s_k}^{\epsilon_k} g_{t_1}^{\delta_1} \dots g_{t_l}^{\delta_l}$$

and on the other hand

$$H([w])H([w']) = g_{s_1}^{\epsilon_1} \dots g_{s_k}^{\epsilon_k} g_{t_1}^{\delta_1} \dots g_{t_l}^{\delta_l}$$

□

Using this we can show that any group free on a subset of its elements in fact looks like one of the formal groups that we constructed.

Proposition 1.29: *Suppose G is free on S, then G is isomorphic to $F(S)$.*

Proof. Consider the homomorphism $F(S) \to G$ given by extending the identity map $S \to S$. It is surjective, since S generates G. It is injective: pick an element in the kernel, take its reduced representative $w = s_1^{\epsilon_1} \dots s_k^{\epsilon_k}$, we have $H([w]) = s_1^{\epsilon_1} \dots s_k^{\epsilon_k} = 1$. Since G is free on S and w is reduced it must be the empty word. Thus H is also injective. □

We also prove

Lemma 1.30: *Let S, S' be sets. Then $F(S)$ is isomorphic to $F(S')$ iff $|S| = |S'|$.*

Thus the isomorphism type of G depends only on the cardinality of S. When it is finite of cardinality k, we denote G by $\mathbb{F}_k$.

Note that however free groups on sets of different cardinalities are distinct.

Proof. Suppose $|S| = |S'|$. Take h extending $S \to S'$ bijection and h' extending $S' \to S$ inverse bijection, then $h' \circ h$ is a group morphism $F(S) \to F(S)$ which extends the identity on S: by the Lemma this must be the identity (using uniqueness), hence h and h' are isomorphisms.

Other direction: count the number of morphisms $G \to \mathbb{Z}/2$: each of the $2^{|S|}$ choices of image for S gives a unique morphism by universal property, and every morphism is obtained in this way. Thus $2^{|S|}$ morphisms. If G isomorphic to G', there are exactly as many morphisms $G' \to \mathbb{Z}/2$ as $G \to \mathbb{Z}/2$, hence $|S| = |S'|$ (for infinite cardinals this requires the generalized continuum hypothesis or remark that card of G equals that of $|S|$). □

Note that if G is free on S the set S is generating, and there are no non-trivial relations between the elements of S - by a natural analogy with vector spaces we give

Definition 1.31: *If G is free on S, we say S is a* ***basis*** *for G. We call $|S|$ the rank of the free group (note that it is independent of the choice of the basis by the previous proposition).*

From now on we will drop the formalism of equivalence classes and think of elements in the free group over S as reduced words on S.

So far two characterizations of free groups: either group with a generating set on which there are no nontrivial relations, or the abstract group of (equivalence classes of) words on a set S with concatenation. But in fact the lemma above gives a third characterization, via a "universal property".

Proposition 1.32: *A group G is free on $S \subseteq G$ iff for any group H, any map $h : S \to H$ admits a unique extension to a morphism $G \to H$.*

Proof. Assume that every map $S \to H$ to a group H, admits a unique extension $G \to H$. Consider the identity map $S \to F(S)$, and its unique morphism extension $i : G \to F(S)$. On the other hand, $F(S)$ satisfies the universal property, so there is a unique extension $j : F(S) \to G$ to the identity $S \to G$. This gives a homomorphism $i \circ j : F(S) \to F(S)$ extending $S \to S$: again by universal property of $F(S)$, there is a unique such morphism, the identity. Thus G isomorphic to $F(S)$ via i: deduce from this that G is free on S. □

Again this is consistent with the use of the term "basis" - in a vector space V, to specify a linear map from V, it is enough to specify the image of a basis, and the map is uniquely determined.

Exercise 1.33: Show that if u is a proper power in $F(S)$, it cannot be part of a basis.

Some things to be careful about with free groups:

Remark 1.34:

1. *Not every group admits a basis, only free groups!*
2. *Not every generating set contains a basis.*
3. *A free group admits lots of bases, for example $\{a, ba\}$ is a basis for $\mathbb{F}(a, b)$ (show that the morphisms $a \mapsto a, a \mapsto ba$ and $a \mapsto a, a \mapsto ba^{-1}$ are inverses, hence isomorphisms).*
4. *If S is a subset of G which is free (i.e. no nontrivial reduced word on S represents the identity element) but not generating, it cannot in general be extended to a basis of G.*
5. *A free group of rank k may have free subgroups of rank $n > k$, indeed of infinite rank!*

Exercise 1.35: Consider the subgroup H of $\mathbb{F}(a, b)$ generated by $S = \{h_n = b^n ab^{-n}, n \in \mathbb{N}\}$, and show it is free on S.

Clearly here S cannot be extended to a basis of $F(a, b)$!

1.4 Group presentations

So drawing the Cayley graph $X(G, S)$ of a group G with respect to S, we see relations between elements of S as loops. In general there are non trivial relations as well. Example of $\mathbb{Z}^2$: we have $aba^{-1}b^{-1} = 1$ but also $b(aba^{-1}b^{-1})b^{-1}$ and $ab^2a^{-1}b^{-2}$ etc. But in some sense we do not need to give explicitly these other relations, they are consequences of the first one. If r is a relation on S then so is wrw^{-1} for any word w, and if r_1, r_2 are relations on S then so is $r_1 r_2$.

What are all the relations on S in G? They are the (reduced) words on S which represent the identity element in G. Consider the free group $F(S)$ on S (as built above) and the unique morphism $F(S) \to G$ extending the identity. It is surjective since S generates G. Each word is sent to the element it represents in G. Thus elements of the kernel are precisely relations between the generators.

Definition 1.36: *Let G be a group and S a generating set. Let $R \subseteq F(S)$. Denote by π the morphism $F(S) \to G$ extending the identity on S. We say that G admits the presentation $\langle S \mid R \rangle$ if R normally generates* Ker π*, that is, if* Ker π *is the smallest normal subgroup containing R (we denote this subgroup by $\langle\langle R \rangle\rangle$).*

Exercise 1.37: Show that an element $g \in F(S)$ is in the subgroup normally generated by R iff $g = \Pi_{j=1}^{m} u_j r_j u_j^{-1}$ for elements $u_1, \ldots, u_m$ of $F(S)$ and $r_j \in R \cup R^{-1}$ for all j.

Remark 1.38: *We can also build a group with any presentation we choose: given a set S and a set R of words in S, the quotient $F(S)/\langle\langle R\rangle\rangle$ obviously admits the presentation $\langle S \mid R\rangle$. Note that this is a way to specify a group algebraically, not geometrically.*

Example 1.39: 1. $\langle a, b \mid \rangle$ is the free group on a, b;

2. $\langle a, b \mid aba^{-1}b^{-1}\rangle$ is $\mathbb{Z}^2$;

The idea is that by giving a presentation of a group, we can specify unambiguously a group in a compact form, which should be handy for computations. They are a great tool, but it is not always easy to "recognize" a group from the presentation.

Exercise 1.40: "Recognize" the groups given by the following presentations

- $\langle a, b, c \mid a^2cb^{-1} = aba^{-1}b^{-1}\rangle$
- $\langle a, b \mid aba^{-1}b^{-1}; a^4 = 1\rangle$
- $\langle a, b, c \mid a^2, b^2, c^2, abc\rangle$
- $\langle r, s \mid r^6; s^2; srs^{-1} = r^{-1}\rangle$

In fact it is not even easy to see from the presentation basic things like: is the group finite? The famous Burnside problem is an example: in 1902, Burnside asked whether there could be a finitely generated group which is infinite but in which any element has a finite order. (If drop the finitely generated requirement, take $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z} \times \ldots$). An important variation is the bounded Burnside problem, which asks whether a finitely generated group in which all the elements have bounded order is necessarily finite. It is not too hard to see that this problem is equivalent to the following question

Question 1: *Consider the group*

$$B_{n,m} = \langle a_1, \ldots, a_n \mid w(a_1, \ldots, a_n)^m = 1\rangle$$

Is it infinite?

This is the Burnside problem (1902) - Burnside proved finiteness for $m = 2, 3$, Sanov for $m = 4$, Marshall Hall for $m = 6$.

Novikov and Adian (1968) showed it is infinite for $n \geq 2$ and m odd ≥ 667.

Exercise 1.41: Show that the group $\langle a, b \mid w(a, b)^2 = 1$ for all words $w\rangle$ is finite. [Hint: show first that it is abelian].

1.5 Morphism from a presentation

We saw that any choice of images in a target group G for the generators of a free group $F(S)$ extend to a morphism. If the group is not free, but admits a presentation $\langle S \mid R\rangle$, how do we need to choose our images to make sure this extends to a morphism?

Proposition 1.42: *Let G be the group $\langle S \mid R\rangle$. Let H be any group.*

Given any map $S \to H$ given by $s \mapsto \bar{s}$, there exists a morphism $G \to H$ extending it if and only if the relations of R are satisfied by the elements $\bar{s}$, i.e. if for any word $r(s_1, \ldots, s_m)$ in R, the product $r_H(\bar{s}_1, \ldots, \bar{s}_m)$ is trivial in H.

Notation: if $w(s_1, \ldots, s_m)$ is a reduced word $s_{i_1}^{\epsilon_1} \ldots s_{i_k}^{\epsilon_k}$ in $F(S)$, and if $g_1, \ldots, g_m$ are elements in a group G, we denote by $w_G(g_1, \ldots, g_m)$ the element of G given by the product $g_{i_1}^{\epsilon_1} \ldots g_{i_k}^{\epsilon_k}$.

Proof. Denote by π the morphism $F(S) \to G$ extending $S \to G$.

Suppose there exists a morphism $\theta : G \to H$ extending $s \mapsto \bar{s}$. Then

$$r_H(\bar{s}_1, \ldots, \bar{s}_m) = \theta(r_G(s_1, \ldots, s_m)) = \theta(1) = 1$$

since θ is a morphism.

For the other direction, let $\phi : F(S) \to H$ be the unique morphism extending $s \mapsto \bar{s}$. For any $r(s_1, \ldots, s_m) \in R$, we have that $\phi(r(s_1, \ldots, s_m)) = r_H(\bar{s}_1, \ldots, \bar{s}_m) = 1$ by hypothesis. Thus Kerϕ contains R, and it is normal - it must thus contain $\langle\langle R \rangle\rangle =$ Kerπ. Hence ϕ factors through π i.e. $\phi = \phi' \circ \pi$. The morphism $\phi' : G \to H$ extends $s \mapsto \bar{s}$. $\square$

Exercise 1.43: Let $G = \langle a_1, b_1, a_2, b_2, c \mid c = w(a_1, b_1) = v(a_2, b_2) \rangle$. Show that there exists an isomorphism $\tau : G \to G$ with $\tau(a_1) = a_1, \tau(b_1) = b_1, \tau(a_2) = ca_2c^{-1}, \tau(b_2) = cb_2c^{-1}$.

1.6 Geometric viewpoint - Van Kampen diagrams

Geometric viewpoint: relations are loops in the Cayley graph based at 1. What does the conjugate of a relation looks like? The product of two relations?

See on the examples above in $\mathbb{Z}^2$. A big loop can be "paved" by smaller loops. A presentation is a way to give enough small loops so as to be able to recover all of the big ones.

We formalize this idea with Van Kampen diagrams.

Definition 1.44: *(diagram) A diagram M is (the topological realization of a) finite connected planar graph M. We think of edge as doubles, each edge comes with an orientation and with its oppositely oriented edge. Note that the bounded components of $\mathbb{R}^2 - M$ are homeomorphic to the open unit disk.*

Denote by $\hat{M}$ the union of M with the bounded components of $\mathbb{R}^2 - M$.

A boundary cycle of M is a cycle of minimal length which contains all the edges in the boundary of M (there might be several).

Definition 1.45: *(Van Kampen diagrams) If $\langle S \mid R \rangle$ is a presentation (we assume the words in R to be reduced), a Van Kampen diagram associated to $\mathcal{P}$ is a diagram with*

1. *a basepoint v_0;*
2. *a labelling of the edges by elements of S, such that if e is labelled by s then $\bar{e}$ labelled by s^{-1};*
3. *for each region, there is a choice of starting vertex an orientation for the boundary such that the boundary label is in R.*

If the differents words labelling boundary cycles starting at v_0 are reduced, we say that the diagram is reduced.

Remark 1.46: *If M is a Van Kampen diagram associated to the presentation $\langle S \mid R \rangle$ of a group G, there is a morphism of oriented labelled graphs from M to the Cayley graph $X(G, S)$. If M is reduced, the morphism is locally injective.*

Lemma 1.47: *A word w in the letters A represents the identity in $G = \langle A \mid R \rangle$ iff there exists a reduced Van Kampen diagram whose boundary is labelled by w.*

We show the "only if" on an example

Example 1.48: Take the presentation $\langle x, y, t \mid xt^2, x^2ty \rangle$, and consider the relation $(xt^2)t^{-2}(x^2ty)t^2 = x^3tyt^2$.

In general: if w is a relation, can write it as a product of conjugates of elements of R - draw the diagram with a balloon for each factor, and fold until there is nothing left to fold.

1.7 Presentations as a computing tool - Dehn's algorithmic questions

If we want to use presentations to compute things algorithmically, we need them to have some finiteness properties.

Definition 1.49: *We say that a group is finitely generated if it admits a presentation $\langle S \mid R \rangle$ where S is finite, and finitely presented if it admits a presentation $\langle S \mid R \rangle$ where both S and R are finite.*

Remark 1.50: *A finitely generated group is countable. Is every countable group finitely generated?*

Exercise 1.51: Show that there are countable non finitely generated groups.

1. Let S be an infinite set, and consider the free group $F(S)$: suppose it admits a finite generating set U. Each U is a word in finitely many of the letters s, so there is a finite subset S_0 of S such that the subgroup generated by U is contained in the subgroup generated by S_0. But let $s \in S - S_0$: this means s can be written as a product of elements of S_0 and gives a non trivial relation between elements of S.

2. $(\mathbb{Q}, +)$ is countable, but not fg as a group: if $(p_1/q_1, \ldots, p_k/q_k)$ were a generating set, any element in $(q_1 \ldots q_k)\mathbb{Q}$ would be an integer - not the case.

Question 2: *Does every countable group embed in a fg group? in a fp group?*

Question 3: *Are there finitely generated groups which are not finitely presented?*

We will see how to answer this in the sequel...

Even if we restrict ourselves to finite presentations, it is not clear that we can answer any question algorithmically. In 1912, Dehn asked the following questions:

Question 4: *(Word problem) Is there an algorithm which, when given as an input a finite presentation $\langle s_1, \ldots, s_m \mid r_1, \ldots, r_l \rangle$ for a group G, and a word $w(s_1, \ldots, s_m)$, decides whether w represents the identity in G, i.e. whether or not $w_G(s_1, \ldots, s_m) = 1$?*

Remark 1.52: *This is equivalent to deciding whether two words represent the same element.*

Example 1.53: The word problem is solvable in the free group given by the free presentation. Idea: there is an algorithm which "reduces" a word.

Remark 1.54: *It is easy to find an algorithm which given a presentation $\langle S \mid R \rangle$ for a group G and a word $w(s_1, \ldots, s_m)$, terminates if and only if $w_G(s_1, \ldots, s_m) = 1$. Indeed, we saw that $w_G(s_1, \ldots, s_m) = 1$ iff $w(s_1, \ldots, s_m) \in \langle\langle R \rangle\rangle$ iff there exist an integer k, elements $u_i \in F(S)$ and $r_i \in R \cup R^{-1}$ such that the following equality holds in $F(S)$*

$$w(s_1, \ldots, s_m) = \Pi_{i=1}^k u_i(S) r_i^{\pm 1} u_i^{-1}(S).$$

We can thus build an algorithm which enumerates all the reduced words in $\langle\langle R \rangle\rangle$: start by enumerating all the (finitely many) products with $k \leq 1$ and $l_S(u_j) \leq 1$, then those with $k \leq 2$ and $l(u_j) \leq 2$, etc (compute the product as a word, then reduce the word). If the (reduced) word w shows up, stop.

A variant on the word problem is the following

Question 5: *(Conjugacy problem) Is there an algorithm which, when given as an input a finite presentation $\langle s_1, \ldots, s_m \mid r_1, \ldots, r_l \rangle$ for a group G, and two words $w(s_1, \ldots, s_m)$ and $u(s_1, \ldots, s_m)$, decides whether w and u represent conjugate elements of G?*

Remark 1.55: *(Word problem is easier than the conjugacy problem) If a group has solvable conjugacy problem, it has solvable word problem (deciding whether a word is trivial is the same as deciding whether it is conjugate to the identity.)*

There are many possibilities for the choice of R, i.e. a group admits many different presentations. One can also choose a different generating set. This led Dehn to ask the following question

Question 6: *(Isomorphism problem) Is there an algorithm which, when given as an input two finite presentations $\langle s_1, \ldots, s_m \mid r_1, \ldots, r_l \rangle$ and $\langle t_1, \ldots, t_k \mid u_1, \ldots, u_j \rangle$, decides whether the groups given by these presentations are isomorphic?*

Remark 1.56: *If we remove the "finitely presented" condition, of course there are some groups with unsolvable word problem: take S to be a non recursive subset of $\mathbb{N}$ (see Definition 1.77) and look at: $\langle a_1, a_2, \ldots \mid a_i$ for i in $S \rangle$ How can we make this finitely generated? presented?*

We will see that the answers to these questions are negative - however, it is still interesting to know for which classes of groups these hold.

1.8 Building new groups from old ones - amalgamated products and HNN extensions

Example 1.57: Direct product, semidirect product, quotients

In both direct and semidirect product, there are "relations" between the elements of the 2 groups we wish to combine.

Definition 1.58: *Let A, B be two subgroups of a group G such that*

1. *A, B generate G;*
2. *no product $a_1 b_1 \dots a_r b_r$ where $a_i \in A - \{1\}$ for all j and $b_j \in B - \{1\}$ for all $j < r$ is trivial.*

Then we say that G is the free product of A and B, and we write $G = A * B$.

Remark 1.59: *The second condition is equivalent to*

(ii)' no product $d_1 d_2 \dots d_r$ where $d_i \in A \cup B - \{1\}$ and $d_i \in A$ iff $d_{i+1} \in B$ for all $i < r$ is trivial.

(we sometimes call such a product an "alternating product"). Indeed, clearly the product in (ii) is alternating so $(ii)' \to (ii)$. To see the inverse direction, note that if $d_1 d_2 \dots d_r$ is an alternating product with $d_1 \in B$, we can pick $a \in A - \{1\}$ and condider the product $a d_1 \dots d_r a^{-1}$. It is a product like in (ii), hence it is not trivial. But this implies that $d_1 \dots d_r$ is non trivial.

Example 1.60: The free group on $\{a, b\}$ is the free product of $A = \langle a \rangle \simeq \mathbb{Z}$ with $B = \langle b \rangle \simeq \mathbb{Z}$.

Remark 1.61: *If g is a finite order element in $A * B$, then some conjugate of g lies in A or in B - indeed, up to conjugation we can write any element which does not satisfy this requirement as $g = a_1 b_1 \dots a_r b_r$ with $a_1 \neq 1$ and $b_r \neq 1$, and then we see that no power of g is trivial.*

Constructive approach:

Proposition 1.62: *Given two groups A and B, pick presentations $\langle S_A \mid R_A \rangle$ and $\langle S_B \mid R_B \rangle$ respectively for A and B. Let G be the group given by the presentation $\langle S_A, S_B \mid R_A, R_B \rangle$. The groups generated by S_A, S_B in G are (isomorphic to) A, B respectively, and G is the free product of these two subgroups.*

Exercise 1.63: Show that if $G = A * B$, we also have $G = A * aBa^{-1}$ for any $a \in A$. Is this true if replace a by any g in G?

Example 1.64:

1. Infinite dihedral group $\mathbb{Z}/2\mathbb{Z} * \mathbb{Z}/2\mathbb{Z}$
2. For those who know π_1: wedge sum of two space (Van Kampen)

Now suppose we want to create a group out of A and B but that we want to identify two subgroups, in A and B respectively, which are isomorphic.

Definition 1.65: *Let $A = \langle S_A \mid R_A \rangle$, $B = \langle S_B \mid R_B \rangle$, C a group with embeddings $C \to A, c \mapsto c_A$, $C \to B, c \mapsto c_B$. The amalgamated product of A with B over C is the quotient of $A * B$ by the normal subgroup generated by the elements $c_A c_B^{-1}$ for all $c \in C$. It is denoted by $A *_C B$ (abuse of notation).*

If $\langle S_A \mid R_A \rangle$ and $\langle S_A \mid R_A \rangle$ are presentations of A and B respectively, we can write each element c_A (resp c_B) as a word $c_A(S_A)$ (respectively $c_B(S_B)$) in the letters S_A (respectively S_B). The amalgamated product $A *_{C_A = C_B} B$ is given by

$$\langle S_A \cup S_B \mid R_A \cup R_B \cup \{c_A(S_A) = c_B(S_B) \text{ for every element } c \text{ of } C\} \rangle$$

Another related construction is the HNN extension: given a group with two isomorphic subgroups, add an element which "forces them" to be conjugate.

Definition 1.66: *Let A and C be groups, and suppose there are two embeddings $C \to A, c \mapsto c_1$, $C \to A, c \mapsto c_2$.*

The HNN extension of A over C is the quotient of the free product $A * \langle t \rangle$ by the subgroup normally generated by elements of the form $t c_1 t^{-1} c_2^{-1}$. It is denoted by $A *_C$ (abuse of notation) or by $A *_{C1 = C_2}$

Idea: we have two copies of C, and we want to force them to be conjugate. Problem - does A embed in G?

Lemma 1.67: *The surjective morphism $A * \langle t \rangle \to A*_C$ is injective on A.*

Proof. We consider the full presentation $A = \langle t, A - \{1\} \mid R_A \rangle$. The group G admits the presentation $\langle a \in A - \{1\} \mid R_A, tct^{-1} = \phi(c)$ for all $c \in C \rangle$.

We want to show that if we think of an element $a \in A - \{1\}$ as a (one letter) word in the generators of this presentation, it does not represent the identity element in G. Choose a minimal counterexample in the following way: among all the elements $a \in A - \{1\}$ so that the word $w = a$ represents the identity element in G, choose one which admits a reduced Van Kampen diagram M with minimal number of regions.

If there are no edges labeled by t, all regions are labeled by (conjugates of) relations from R_A, so M is in fact a Van Kampen diagram for the presentation $\langle A - \{1\} \mid R_A \rangle$ of A. Hence a represents the trivial element already in A - this is a contradiction. Let thus e be an edge of M labeled by t. Since e does not lie on the boundary, there are two possibilities: 1. two regions with labels $tc_1t^{-1}\phi(c_1), tc_1t^{-1}\phi(c_1)$ share e. But then we can replace them by a single region also labeled by a relation. 2. one region labeled by $tct^{-1}\phi(c)$ has its two t-edges identified. But then we get that c or $\phi(c)$ admits a Van Kampen diagram with strictly less regions than M, a contradiction. Thus no element $a \in A - \{1\}$ is such that a represents the identity element in G. □

The following is a generalization of this lemma. It says essentially that if a word in the generators $A - \{1\} \cup \{t\}$ of the HNN represents the identity, it means that one of the "relations" appears as a subword, more correctly, a subword of the form tc_1t^{-1} or $t^{-1}c_2t$.

Lemma 1.68: *(Britton's lemma) Let $G = A*_C = \langle A, t \mid tat^{-1} = \phi(a) \rangle$ be an HNN extension. Let w be an element of $A * \langle t \rangle$ given by $w = a_0t^{\epsilon_1}a_1t^{\epsilon_2} \dots t^{\epsilon_n}a_n$ with $\epsilon_i = \pm 1$ and a_i are nontrivial elements of A.*

If w represents the identity in G, then $n \geq 1$ and there is i such that

(i) either $\epsilon_i = 1, \epsilon_{i+1} = -1$, and $a_i \in C$,

(ii) or $\epsilon_i = -1, \epsilon_{i+1} = +1$, and $a_i \in \phi(C)$.

To prove Britton's Lemma, we will use Van Kampen diagrams (the proof is taken from Miller III and Schupp, The geometry of HNN extensions).

Proof. (of Britton's Lemma) We consider the full presentation $A = \langle t, A - \{1\} \mid R_A \rangle$. The group G admits the presentation $\langle a \in A - \{1\} \mid R_A, tct^{-1} = \phi(c)$ for all $c \in C \rangle$.

The previous Lemma shows that we must have $n \geq 1$.

Now suppose $w = a_0t^{\epsilon_1}a_1t^{\epsilon_2} \dots t^{\epsilon_n}a_n$ represents the identity in G. Choose a reduced Van Kampen diagram with boundary label w with minimal number of regions.

Using the same argument as in the proof of the previous lemma, we can show that all the edges labeled by t must lie on the boundary of M. Indeed, if e is an interior edge labeled by t, either it belongs to two different regions which can be identified to produce a smaller diagram, or there is a region labeled by $tc_1t^{-1}c_2^{-1}$ which has its two t-edges identified, but then we get that c_1 or c_2 admits a Van Kampen diagram which contradicts the proof of the previous lemma.

Now t edges come by pairs (each pair corresponding to a region), and pairs cannot "cross" - choose an innermost pair corresponding to a region labeled by $tc_1t^{-1}c_2^{-1}$. We must have that the word labeling the boundary subpath of M between the two edges represents the same element as c_1 or as c_2 in A, hence it also lies in C_1 or in C_2. This proves the claim. □

Corollary 1.69: *If $f \in A*_C$ is a finite order element, then f lies in a conjugate of A.*

Proof. We may assume f non trivial. Suppose f can be written as $a_0t^{k_1}a_1 \dots t^{k_r}a_r$ with $r > 0$ (i.e. that t appears at least once).

Up to conjugating f, may assume that $a_r = 1$ (here use $r > 0$), that if $a_0 = 1$, then $k_1 k_r > 0$, that if a_0 in C_1 then $k_1 < 0$ or $k_r > 0$, and that if $a_0 \in C_2$, then $k_1 > 0$ or $k_r < 0$.

Because f^k is trivial, Britton's lemma tells us that either

- $a_0 = 1$ but in this case we know $k_1 k_r > 0$ and the word f^k cannot represent the trivial element;
- a_0 is in C_1 and $k_r > 0$ and $k_1 < 0$ (but we ensured this cannot be);
- a_0 is in C_2 and $k_r < 0$ and $k_1 > 0$ (but we ensured this cannot be).

□

1.9 Some answers to our questions

We saw that not every countable group is fg (free group on infinitely many elements, $\mathbb{Z}/2\mathbb{Z}^{(\mathbb{N})}$). However, we get a compromise:

Proposition 1.70: *Every countable group embeds in an fg group*

Proof. Let $a_0, a_1, a_2, \ldots$ enumerate the non trivial elements of the group G. Consider the group $A = G * \langle s \rangle$: it is generated by $\{s, a_0, a_1, \ldots\}$. Its two subgroups H_1 generated by $S_1 = \{a_0, sa_1s, s^2a_2s^2, \ldots\}$ and H_2 generated by $S_2 = \{sa_1s, s^2a_2s^2, \ldots\}$ respectively are isomorphic: indeed, can show H_i is free on S_i.

Build the HNN extension $A *_{H_1=H_2}$: it is the quotient of $A * \langle t \rangle$ by the relations $ts^ia_is^it^{-1} = s^{i+1}a_{i+1}s^{(i+1)}$, so it is generated by $\{t, s, a_0\}$ (have $a_1 = s^{-1}ta_0t^{-1}s^{-1}$, then $a_2 = s^{-2}ta_1st^{-1}s^{-2}$, etc.)

By Britton's lemma, A embeds in this group, hence so does G. □

What about embedding in an FP group? The following two propositions shows this is not possible:

Proposition 1.71: *There are uncountably many finitely generated groups.*

Proof. For each infinite subset P of the set $\mathcal{P}$ of prime numbers, define the group $Z_P = \bigoplus_{p \in P} \mathbb{Z}/p\mathbb{Z}$. Each of these subgroups can be embedded in an fg group G_P using the method above (HNN extension of $Z_P * \langle s \rangle$ over some free subgroups). Now if P and Q are distinct subsets of the primes, G_P and G_Q are not isomorphic: indeed, let $p \in P - Q$: G_P contains a subgroup isomorphic to $\mathbb{Z}/p\mathbb{Z}$, while if G_Q contained such a subgroup, it would have to be a conjugate of a subgroup of $Z_Q * \langle s \rangle$, so by Remark 1.61 it would have to be a subgroup of Z_Q - not the case. □

Remark 1.72: *There are up to isomorphism countably many finitely presented groups. Indeed, if fix a max number n of generators, and a max number and size r of relations, there are only finitely many possibilities.*

Note that this shows in particular that there exist finitely generated groups which are not finitely presented. In fact, we have:

Remark 1.73: *There are up to isomorphism countably many finitely generated subgroups of finitely presented groups. Indeed, given a fp group G, fix a finite presentation of it. Count its fg subgroups: if fix a max on number of generators for a subgroup H, and a max for the length of these generators as words in the generators of G, there are only finitely many possibilities for H.*

Thus more generally, there exist fg groups which do not embed in fp groups.

In fact Higman proved in 1961:

Proposition 1.74: *A group G embeds in an fp group H if and only if it is recursively presented (i.e. it has a presentation with finitely many generators and a set of relators which is recursively enumerable). Moreover, if G is torsion free and recursively presented, we may assume H is also torsion free.*

Recursively enumerable - that can be enumerated by an algorithm.

A note on the proof of Higman Theorem: it is fairly easy to see that any fg subgroup of a fp group is recursively presented, the other direction is much harder. The group H is constructed from G by a sequence of HNN extensions, hence Corollary 1.69 implies the "moreover" part of the result.

What is "an algorithm"? We need a mathematical model for this.

An algorithm is a finite set of instructions that can be implemented by a Turing machine.

Given an input $(a_1, \dots, a_k)$, a Turing machine will apply the set of instructions. Two things can happen: either 1. after finitely many steps, it reaches its "stopping state" and produces an output $f(a_1, \dots, a_k)$ which is a positive natural number or 2. it never reaches the stopping state. In other words, Turing machines compute functions defined on subsets of $\mathbb{N}^k$ to $\mathbb{N}$ (never stops iff the function was not defined there).

Definition 1.75: *A function which can be computed by a Turing machine is called computable, or equivalently, recursive.*

Equivalently we can give

Definition 1.76: *The class of recursive (partial) functions is the smallest class of functions containing : the constant functions, the projection-to-a-coordinate functions, the successor function; and closed under: composition, recurrence ($f(a+1) = g(f(a), a)$) and the μ "search operator" ($\mu(f)(a) = \min\{b \mid g(b, a) = 0$).*

BUT: not all functions are recursive!

Proof idea: Can encode Turing machines by a number, i.e. can associate uniquely to each machine (and thus to each recursive function f) a natural number e - conversely denote by f_e the function computed by the TM encoded by e. Remark: we can do this in such a way that the function U which to a pair (e, x) associates $f_e(x)$ is recursive (universal Turing machine).

Now consider the function H such that $H(e)$ is 0 if $f_e(e)$ is undefined, and is undefined otherwise. Suppose H were recursive: it would be computed by a TM, let u be this TM's code. Consider now $H(u)$: if $f_u(u)$ is undefined, it is defined and equals 0, and if $f_u(u)$ is defined, it is undefined. But $f_u(u) = H(u)$ so this yields a contradiction.

Definition 1.77: *Let $D \subseteq \mathbb{N}$. We say D is recursive if the characteristic function of D is recursive.*

We say D is recursively enumerable if the function I_D given by $I_D(n) = 1$ if $n \in D$ and $I_D(n)$ undefined otherwise is recursive.

The proof above shows that the set

$$\{e \mid f_e(e) \text{ is defined }\}$$

is not recursive.

Note that its complement $\mathbb{N} - A$ is recursively enumerable (apply the universal Turing machine to $U(e, e)$, if it stops, output 1). Thus we have produced a set which is recursively enumerable but not recursive.

Remark 1.78: *If S is an alphabet, and W is the set of words on S, we can order the words by lexicographic order (fix an order on $S \cup S^{-1}$), this gives a bijection $W \to \mathbb{N}$. Then we say that a subset D of W is recursive (respectively recursively enumerable) if $f(D)$ is. Easy to see this does not depend on choice of ordering we started with.*

Undecidability of the word problem.

Proposition 1.79: *There exists a finitely generated torsion free group with undecidable word problem.*

(the fact that it is torsion free will be useful for the proof of Theorem 1.82 below).

Proof. First, find a non FG example: let S be non recursive. $G = \langle t, a_1, a_2, \dots \mid t a_i t^{-1} = a_i \forall i \in S \rangle$.

It is an HNN extension of G along the subgroup generated by a_i's with $i \in S$. By Britton's lemma, the word $t a_i t^{-1} a_i^{-1}$ is trivial iff $i \in S$, thus there cannot be an algorithm deciding whether this is true.

To get FG group, take: $G = \langle t, a, b \mid tb^i ab^{-i} t^{-1} = b^i ab^{-i} \forall i \in S \rangle$. It is an HNN extension of $\mathbb{F}(a, b)$ hence it is torsion free by Corollary 1.69. $\square$

Exercise 1.80: Show that if H is a finitely generated subgroup of G, and H has unsolvable word problem, then G has unsolvable word problem.

Corollary 1.81: *There exists a finitely presented torsion free group with undecidable word problem.*

Proof. In proof above, if S recursively enumerable but not recursive, by Higman's theorem we can in fact embed G in a torsion free fp group G' which will thus also have unsolvable word problem by the exercise above. $\square$

Proposition 1.82: *The isomorphism problem is unsolvable in general.*

Proof. Suppose $G = \langle A \mid R \rangle$ with $A = \{a_1, \dots, a_m\}$ is a torsion free FP group with unsolvable word problem.

Let $H_w = \langle A, s, t \mid R, t(s^i a_i s^{-i})t^{-1} = s^i w s^{-i} \rangle$. If $w = 1$ in G, then $a_i = 1$ in H_w so H_w is the free group on s, t.

If w is not equal to 1, H_w is an HNN extension of $G* < s >$ with stable letter t where you make the subgroups $< s^i a_i s^{-i} >$ and $< s^i w s^{-i} >$ (which are both free of rank n - thanks to the fact that G is torsion free) conjugate by t, thus it contains a copy of G, thus it has unsolvable word problem, thus it is not free.

Hence H_w is isomorphic to $\mathbb{F}_2$ iff $w = 1$ in G. $\square$

Classes of groups for which this is solvable: finite, abelian, polycyclic-by-finite, nilpotent, torsion free hyp, torsion-free toral relatively hyperbolic, limit groups

Not solvable for: solvable of derived length 3.

2 Quasi-isometry

Problem of Cayley graph: depends on the choice of the generating set. Example of $\mathbb{Z}$: the graph still looks the same viewed from far away. Notion of "coarse geometry". Careful quasi-isometric Cayley graphs does not imply isomorphic groups.

What properties do quasi-isometric groups have in common? Quasi-isometry invariants.

2.1 Quasi-isometric embeddings

Definition 2.1: *Let (X, d_X) and (Y, d_Y) be metric spaces, and $C \geq 1, D > 0$. A map $f : X \to Y$ is a (C, D)-quasi-isometric embedding if for all x_1, x_2 in X we have*

$$\frac{1}{C} d_X(x_1, x_2) - D \leq d_Y(f(x_1), f(x_2)) \leq C d_X(x_1, x_2) + D$$

Careful! A quasi-isometric embedding need not be an embedding...also, it is not necessarily continuous.

The constant are not so important.

Remark 2.2: *There is some symmetry in the definition between $d(f(x_1), f(x_2))$ and $d(x_1, x_2)$. Equivalent formulation: $d_Y(f(x_1), f(x_2)) \leq C d_X(x_1, x_2) + D$ AND $d_X(x_1, x_2) \leq C d_Y(f(x_1), f(x_2)) + CD$.*

Exercise 2.3: The following are quasi-isometric embeddings.

- $\mathbb{Z}$ in $\mathbb{R}$
- logarithmic spiral $c : t \mapsto (t\cos(\log t), t\sin(\log t))$.

This one isn't:

- $\mathbb{Z} \to \mathbb{R}$ given by $n \mapsto n^3$.

Definition 2.4: *Let $f : X \to Y$ be a (C, D)-quasi-isometric embedding. Suppose moreover that for any $y \in Y$ there exists x such that*

$$d_Y(f(x), y) \leq D$$

Then we say that f is a quasi-isometry, and that X and Y are quasi-isometric.

Example 2.5:

- $\mathbb{Z} \to \mathbb{R}$;
- $\mathbb{Z}^2 \to \mathbb{R}^2$
- A metric space has finite diameter iff it is QI to a point;
- The inclusion of G with the word metric in $X(G, S)$ is a QI.

Equivalent definition of quasi-isometry

Proposition 2.6: *Two metric spaces X and Y are quasi-isometric iff there exist maps $f : X \to Y$ and $h : Y \to X$ and constants $C, D > 0$ such that for any $x, x' \in X$ and $y, y' \in Y$*

$$d_Y(f(x), f(x')) \leq C d_X(x, x') + D \text{ and } d_X(h(y), h(y')) \leq C d_Y(y, y') + D$$

$$d_X(x, h(f(x))) \leq D \text{ and } d_Y(y, f(h(y))) \leq D.$$

Proof. Suppose X, Y quasi isometric. For any point y of Y, let N_y be the (nonempty) set of points x of X whose image by f is within a distance D_f of y. By axiom of choice, there exists a function $h : Y \to X$ such that $h(y) \in N_y$.

Let $x, x' \in X$. Since f is a (C_f, D_f)-quasi isometric embedding, the first inequality holds for any $C \geq C_f, D \geq D_f$. Let $y, y' \in Y$ and set $x = h(y), x' = h(y')$ - by definition of h we have $d(f(x), y) \leq D_f$

and $d(f(x'), y') \leq D_f$. Since f is a quasi isometric embedding we have that $d(x, x') = d(h(y), h(y')) \leq C_f d(f(x), f(x')) + D_f C_f$ hence

$$d(x, x') = d(h(y), h(y')) \leq C_f(d(y, y') + 2D_f) + D_f C_f = C_f d(y, y)' + 3D_f C_f$$

which gives the second inequality for any $C \geq C_f$ and $D \geq 3D_f C_f$.

Let $x \in X$, if $y = f(x)$ then $h(y)$ is a point $x' \in X$ such that $d(f(x'), f(x)) \leq D_f$, but we also have $d(x, x') \leq C_f d(f(x), f(x')) + C_f D_f$ so $d(x, h(f(x))) = d(x, x') \leq 2C_f D_f$. Let $y \in Y$, by definition $f(h(y))$ is at a distance at most D_f of y.

Taking $C = C_f$ and $D = 3C_f D_f$ yields the result. □

Using this result, it is easy to show that QI is an equivalence relation.

2.2 Quasiisometry of groups

The following proposition shows that quasi-isometry is indeed the right notion to deal with the dependence of the Cayley graph on the generators.

Proposition 2.7: *Let G be a group, and let S_1, S_2 be two finite generating sets for G. The identity is a quasi-isometry between (G, d_{S_1}) and (G, d_{S_2}).*

Proof. Let $f : G \to G$ denote the identity, and let $C_1 = \max\{l_{S_2}(s) \mid s \in S_1\}$ and $C_2 = \max\{l_{S_1}(s) \mid s \in S_2\}$.

It is not hard to see that $l_{S_2}(g) \leq C_1 l_{S_1}(g)$ and similarly for C_2.We have

$$d_{S_2}(f(g), f(g')) = l_{S_2}(g^{-1}g') \leq C_1 l_{S_1}(g^{-1}g') = C_1 d_{S_1}(g, g'),$$

similarly $d_{S_1}(f(g), f(g')) \leq C_2 d_{S_2}(g, g')$.

Since obviously the identity is an inverse of itself, we are done. □

The *isometry class* of a group seen as a metric space via the word metric depends on the choice of generating set, but what the proposition above shows is that the *quasi-isometry* class doesn't.

Note that the assumption that the S_i are finite is important: look at $\mathbb{Z}$ with generating set $\mathbb{N}^+$: this has finite diameter. From now on unless specified otherwise we consider only finite generating sets - and talk about the QI type of G, without specifying the generating set.

Question. What does the QI class of a group tell us about it? How precisely does it define the group?

Up to isomorphism? NO. There are some groups which are QI though they are not isomorphic.

Example 2.8: G is QI to the trivial group iff it is finite. Trivial group = metric space with a single point. Thus G QI to trivial group iff it has finite diameter (see remark in previous Section).

Now if G is finite, clearly it has finite diameter. Conversely, if it has finite diameter it means that $\forall g \in G$ we can write G as a product of at most R elements of $S \cup S^{-1}$. Since S is finite, so is G.

Example 2.9: $\mathbb{Z} \times \mathbb{Z}/2$, D_∞ and $\mathbb{Z}$ are all QI though they are not isomorphic. In fact, $\mathbb{Z} \times \mathbb{Z}/2$ and D_∞ are even isometric.

In fact both examples are an instance of the following results:

Proposition 2.10: *If H is a subgroup of finite index of a fg group G, then H and G are quasi-isometric.*

Corollary 2.11: *Groups which are commensurable (i.e. which admit isomorphic finite index subgroups) are QI.*

Exercise 2.12: Prove this by choosing suitable systems of generators for H and G.

We will see it in the next section as a consequence of a more general result (Svarc-Milnor Lemma).

So one might wonder if up to finite index, the quasi-isometry class of a group determines its isomorphism type. This is not the case in general (not so easy to see! See Remark 8.21 in Bridson Haefliger, or Proposition 30 in de la Harpe), but it works for some cases:

Proposition 2.13: *A fg group QI to $\mathbb{Z}$ is virtually $\mathbb{Z}$, i.e. it contains $\mathbb{Z}$ as a subgroup of finite index.*

Proposition 2.14: *A group whose Cayley graph is quasiisometric to a tree is virtually free.*

Remark - the corresponding result for groups which are QI to a free abelian group is true, but much harder!

2.3 Švarc-Milnor lemma

The Švarc-Milnor lemma is a way to determine the quasi-isometry class of a group. It says that if a group admits a sufficiently nice action on a metric space, then this metric space is quasi isometric to the Cayley graph.

Need some definitions first.

Definition 2.15: *Let X be a metric space. A geodesic segment is the isometric embedding of a segment $[0, a]$ of $\mathbb{R}$ in X, i.e. a map $\gamma : [0, a] \to X$ such that for any $s, t \in [0, a]$ we have*

$$d_X(\gamma(s), \gamma(t)) = |s - t| \,.$$

Geodesic ray: replace $[0, a]$ by $[0, \infty)$. Geodesic line/ geodesic: replace $[0, a]$ by $\mathbb{R}$.

Example 2.16: Geodesic rays in $\mathbb{R}^2$ are exactly segments of lines. Let $\gamma : [0, a] \to \mathbb{R}^2; t \mapsto (x(t), y(t))$ be a geodesic segment.

Wlog we may assume that $\gamma(0) = (0, 0)$ and $\gamma(a) = (0, b)$. We must in fact have $a = d(\gamma(0), \gamma(a)) = b$. In particular $\int_0^a x'(t)\mathrm{dt} = [x(t)]_0^a = a$.

For any $t_0 \in [0, a]$ we have

$$\sqrt{(x'(t_0))^2 + (y'(t_0))^2} = \|\gamma'(t_0)\| = \lim_{t \to t_0} \frac{\|\gamma(t) - \gamma(t_0)\|}{|t - t_0|} = 1$$

hence $\int_0^a \sqrt{(x'(t))^2 + (y'(t))^2}\mathrm{dt} = a$.

We see that we must have $x'(t) = \sqrt{(x'(t))^2 + (y'(t))^2}$ for all t, in particular $y'(t) = 0$ and $x'(t) = 1$ for all t. Hence $y(t) = 0$ and $x(t) = t$ for all t.

Definition 2.17: *A metric space X is said to be geodesic if for any points x, y in X, there exists a geodesic segment $\gamma : [0, a] \to X$ such that $\gamma(0) = x$ and $\gamma(a) = y$.*

Example 2.18: $\mathbb{R}^2$ is geodesic, but $\mathbb{R}^2 - \{0\}$ isn't. If X is the geometric realization of a graph, then X is geodesic.

Definition 2.19: *A metric space X is said to be proper if closed balls (sets of the form $B_R(x_0) = \{x \mid d(x_0, x) \leq R\}$) are compact.*

Example 2.20: $\mathbb{R}^n$ is proper (compact subsets of $\mathbb{R}^n$ are exactly closed and bounded subsets)

When is a graph endowed with the usual metric proper? First let us show the following:

Remark 2.21: *Let X be the geometric realization of a graph. A subgraph A of X is compact iff it is finite. Indeed, if A is infinite, a sequence of midpoint of distinct edges has no convergent subsequence. For the converse, note that any sequence of points in a finite graph has a subsequence which lives in one of its edges.*

Lemma 2.22: *The geometric realization X of a graph is proper iff it is locally finite (that is, every vertex has finite valency)*

Proof. Suppose X has a vertex with infinite valency. The ball of radius 1 around that vertex is not compact so X is not proper. Suppose X is locally finite. Consider a closed ball $B = B_R(x_0)$: the minimal subgraph A containing B is bounded (it is within distance $R+1$ of x_0) and locally finite, hence finite. By the remark above, A is thus compact. Since B is a closed subset of a compact set, it is itself compact. $\square$

We give the following definition in the context of metric spaces, but it can be given for a general topological space.

Definition 2.23: *Let G be a group acting on a metric space X. Say the action is properly discontinuous if for any compact subset K of X, the set*

$$\{g \in G \mid g \cdot K \cap K \neq \emptyset\}$$

is finite.

In particular, point stabilizers are finite - it is a "freeness" condition.

Remark 2.24: *Some textbooks define this by "any point has an open neighborhood U such that the set $\{g \in G \mid g \cdot U \cap U \neq \emptyset\}$ is finite. For locally compact spaces, this is equivalent.*

Also, note that "properly discontinuous" is a special case of a proper action: in general, one also considers the group G to be endowed with a topology, and says that the action is proper if the map $(g, x) \mapsto (gx, x)$ is proper. Properly discontinuous is if moreover the group is endowed with the discrete topology.

It also ensures that the quotient is Hausdorff in the quotient topology.

Example 2.25:
- Any action of a group on a graph where stabilizers of points are finite is properly discontinuous - indeed, let K be compact: wlog we may assume it's a subgraph. The set V_K of its vertices is finite. Now since elements of G send vertices to vertices, we get

$$\{g \in G \mid g \cdot K \cap K \neq \emptyset\} = \bigcup_{v,v' \in V_K} \{g \mid g \cdot v = v'\}$$

Now if $g, h \in \{g \mid g \cdot v = v'\}$, we get $g^{-1}h \cdot v = v$ hence the sets $\{g \mid g \cdot v = v'\}$ are finite. This proves the claim.

- Action of $\mathbb{Z}$ on $\mathbb{R}$, of $\mathbb{Z}^2$ on $\mathbb{R}^2$.
- Action of π_1 of a locally compact path-connected topological space on its universal cover (assuming it admits one). Indeed, the action of $\pi_1(X)$ on $\tilde{X}$ is free and properly discontinuous. (Equivalent definition: it satisfies: for any point $x \in \tilde{X}$, there exists a neighborhood U of x such that for any $g \in G - \{1\}$ we have $U \cap g \cdot U = \emptyset$).
- The action of an infinite group on a compact set cannot be properly discontinuous: for example, the action of $\mathbb{Z}$ on S^1 by an irrational rotation is not properly discontinuous even though it is free.

Exercise 2.26: Show that the subgroup of isometries of $\mathbb{R}$ generated by the translation of length 1 and that of length a for a irrational is free abelian. Show that its action on $\mathbb{R}$ is not properly discontinuous.

Again the following definition can also be given in the wider context of topological spaces.

Definition 2.27: *An action of a group G on a metric space X is said to be cocompact if there exists a compact K such that $X = \bigcup_{g \in G} g \cdot K$.*

Example 2.28:
- The action of a group on a graph is cocompact iff there are finitely many orbits of vertices and edges. For example, the action of a group on its Cayley graph is cocompact (all the vertices are in the same orbit, and there is one orbit of edges for each $s \in S$).
- The actions of $\mathbb{Z}$ on $\mathbb{R}$ (take closed bounded interval of length 1), of $\mathbb{Z}^2$ on $\mathbb{R}^2$ (take square of side 1) are cocompact.

- More generally, the action of the π_1 of a COMPACT locally compact path-connected topological space on its universal cover is cocompact.
- The action of $\mathbb{Z}$ on $\mathbb{R}^2$ is not cocompact.

The following proposition is sometimes called the "fundamental theorem of geometric group theory".

Proposition 2.29: *(Švarc-Milnor lemma) Let X be a proper geodesic metric space. Let G be a group acting on X cocompactly and properly discontinuously by isometries. Then G is finitely generated, and quasi-isometric to X via the orbit map $g \mapsto g \cdot x_0$.*

Proof. Let K compact be such that $X = \bigcup_{g \in G} g \cdot K$. Since K is compact, it has finite diameter, thus it is contained in a closed ball $B = B(x_0, r)$ of radius r around a point x_0.

Let $S = \{g \in G \mid g \cdot B(x_0, 3r) \cap B(x_0, 3r) \neq \emptyset\}$. It is finite by hypothesis of proper discontinuity.

Let $C = \max_{s \in S} d_X(x_0, s \cdot x_0)$.

Let g be in G, consider a geodesic segment γ between x_0 and $g \cdot x_0$, and points $x_1, x_2, \ldots, x_k = g \cdot x_0$ on this segment such that $d(x_{i-1}, x_i) < r$ and k is minimal for this property (i.e. $k = \lceil d(x_0, g \cdot x_0)/r \rceil$).

Each x_i lies in a translate of B. In particular, we can find a sequence $g_0 = 1, g_1, g_2, \ldots, g_k = g$ in G such that $x_i \in g_i \cdot B$. Since

$$d(g_{i-1} \cdot x_0, g_i \cdot x_0) \leq d(g_{i-1} \cdot x_0, x_{i-1}) + d(x_{i-1}, x_i) + d(x_i, g_i \cdot x_0) \leq 3D$$

we get that $g_{i-1}B \cap g_i B$ is non empty, so that $g_{i+1} = g_i s$ for some $s \in S$. Hence g is the product of k elements of S and S generates G.

We have that

$$l_S(g) \leq k \leq \frac{1}{r} d(x_0, g \cdot x_0) + 1.$$

On the other hand, for any sequence $s'_1, \ldots, s'_{k'}$ such that $g = s'_1, \ldots, s'_{k'}$, we get by the triangular inequality that

$$d(x_0, g \cdot x_0) \leq \sum_{i=1}^{k'} d(x_0, s_i \cdot x_0) \leq Ck'$$

hence $d(x_0, g \cdot x_0) \leq Cl_S(g)$.

The map $g \mapsto g \cdot x_0$ is thus a quasi-isometric embedding. Its image is quasi dense since the balls $g \cdot B_R(x_0)$ cover X. This it is a quasi-isometry. ☐

Consequences:

Proposition 2.30: *If G has finite index in H then they are QI.*

Proof. Consider the action of G on $X(H, S)$: it is free (action of H on $X(H, S)$ is free), hence it is properly discontinuous by Example 2.25 above. It is cocompact: let $H = \sqcup_{Gh_i}$, and let x be a vertex of $X(H, S)$. Note that any vertex y of $X(H, S)$ is of the form $h \cdot x$ for some $h \in H$, hence $y = gh_i \cdot x$ for $g \in G$ and some i. Thus any vertex of $X(H, S)$ is in the orbit under G of one of $h_1 \cdot x, \ldots, h_l \cdot x$. Similarly, let $e_1, \ldots, e_r$ be representatives of the orbits of edges under H: any edge of $X(H, S)$ is in the orbit under G of one of the $h_i \cdot e_j$. By the Švarc-Milnor lemma, we get that G is QI to $X(H, S)$, hence to H. ☐

As we saw, this implies that commensurable groups are quasi-isometric, in other words that groups which are isomorphic "up to finite index" are QI. The following proposition shows that if groups are isomorphic "up to a finite kernel", then they are quasi-isometric.

Proposition 2.31: *Let G be a finitely generated group, and let N be a finite normal subgroup of G. Then G is quasi isometric to G/N. If H is a group which admits a finite normal subgroup N' such that $G/N \simeq H/N'$ then G and H are QI.*

Proof. Exercise. ☐

It is possible to show that the free group $\mathbb{F}(a,b)$ admits a free subgroup of rank n as a subgroup of index $n-1$ for all n (we prove this in the sequel). Thus we get

Proposition 2.32: *All the finitely generated free groups of rank ≥ 2 are quasi isometric. In particular, all the regular trees of valency $2k$ are quasiisometric.*

(In fact can show that all regular trees of valency k are quasiisometric.)

Example 2.33: The group $\mathbb{Z}^n$ acts by translations on $\mathbb{R}^n$. The action is properly discontinuous and cocompact. Hence $\mathbb{Z}^n$ is quasiisometric to $\mathbb{R}^n$.

Generalize this in two ways: 1. symmetries of other tilings of the plane; 2. other covering maps.

Example 2.34: Triangle group: group generated by reflections in the sides of an equilateral triangle. Point and edge stabilizers are finite, hence the action is properly discontinuous. Action is cocompact: the closed triangle is a fundamental domain. Hence this group is quasi-isometric to $\mathbb{R}^2$. Presentation of the group: $\langle r,s,t \mid r^2; s^2; t^2; (rs)^3; (st)^3; (rt)^3 \rangle$

In fact can do this for any triangle of angles $\pi/p; \pi/q; \pi/r$ such that $\pi/p+\pi/q+\pi/r=\pi$ (can show that only triple for which this happens are $(3,3,3), (2,2,4)$ and $(6,3,2)$)

In fact, this contains a $\mathbb{Z}^2$ with finite index so that's another way to get this result.

Example 2.35: Let X be a metric space which admits a universal cover $\pi : \tilde{X} \to X$ on which the metric lifts (i.e., there is a metric on $\tilde{X}$ such that $d_X(p,q) = \inf_{p' \in \pi^{-1}(p), q' \in \pi^{-1}(q)} d_{\tilde{X}}(p',q')$). (It is enough for this to be satisfied that the action of $\pi_1(X)$ on $\tilde{X}$ be by isometries)(Example: $\mathbb{R}^2$ covering $\mathbb{T}^2$).

If X is compact, then $\pi_1(X)$ is quasi isometric to $\tilde{X}$.

Example 2.36: Recall definition of manifold: topological space which is locally homeomorphic to $\mathbb{R}^n$ for some n. We want to put a metric on the manifold (inducing the same topology). Many ways to do so.

One way: maybe can show that in fact any neighborhood is isometric to an open subset of $\mathbb{R}^n$ (then need to add some compatibility condition). Then call this a flat metric.

Ex: build the torus as a quotient of the action of $\mathbb{Z}^2$ on $\mathbb{R}^2$ - because the action is by isometries, the metric goes down to the quotient, and we get a flat metric on the torus.

In fact, this is always how flat metrics come to be! Fact: a (connected) topological manifold M of dimension n can be endowed with a flat metric iff its universal cover is $\mathbb{R}^n$, and in this case the covering map is a local isometry.

Another way: if any neighborhood is isometric to an open set in $\mathbb{H}^n$ (with same compatibility condition). Then call this a hyperbolic metric. Fact: a (connected) topological manifold M of dimension n can be endowed with a hyperbolic metric iff its universal cover is $\mathbb{H}^n$, and in this case the covering map is a local isometry.

Let M be a compact and connected manifold. If it admits a flat metric, then $\pi_1(x)$ is quasi isometric to $\mathbb{R}^n$. If it admits a hyperbolic metric, then $\pi_1(M)$ is isometric to $\mathbb{H}^n$

So just by looking at $\pi_1(M)$, can exclude a given topological from admitting a certain Riemannian metric. Also, $\mathbb{R}^n$ and $\mathbb{H}^n$ are not quasi isometric, so we see the topology of the manifold dictates what kinds of metric it can be endowed with.

A much stronger result is Mostow's rigidity theorem, whose proof also uses the Švarc-Milnor Lemma.

Theorem 2.37: *Let M, N be two complete finite volume hyperbolic manifolds of dimension at least 3 (i.e. manifolds endowed with a metric which makes them locally isometric to $\mathbb{H}^n$). If $\pi_1(M)$ and $\pi_1(N)$ are isomorphic, then M and N are isometric.*

2.4 Quasi-isometry invariants

How do we prove that spaces or groups are NOT quasi-isometric? What are quasi isometric invariants that can help us distinguish between quasi-isometry classes? For the moment, we gave only one: the property of having finite diameter is an invariant of QI. In this section, we will present a few more (for a long list, see de la Harpe p115).

Given a quasi-isometry invariant, what does it tell us about the algebraic property of a group?

2.4.1 Ends of a group

Quasi-isometry class = what does the space look like "on a large scale"? For example, what does it look like at infinity?

Definition 2.38: *Number of ends of a graph. Let Γ be a graph, with a fixed vertex v. Let $e_n(\Gamma)$ be the number of unbounded components of $\Gamma - B_v(n)$. It is not hard to see that if $m \geq n$ then $e_m(\Gamma) \geq e_n(\Gamma)$.*

We define the number of ends of Γ to be $e(\Gamma) = \lim_{n\to\infty} e_n(\Gamma)$ - note that it can be infinite.

Remark 2.39: *It is not hard to see that $e(\Gamma)$ does not depend on the choice of v.*

Example 2.40: Finite graph has 0 ends. Comb graph has n ends. Cayley graph of $\mathbb{Z}$ has 2 ends. Cayley graph of $\mathbb{Z}^2$ has one end. Cayley graph of $\mathbb{F}^2$ has infinitely many ends.

Lemma 2.41: *Let X, X' be quasi isometric graphs. Then $e(X) = e(X')$.*

Proof. Let $\phi : X \to X'$ be a (C,D)-quasi isometry, choose a base vertex v in X, and let $v' = \phi(v)$. Wlog can assume there exists a quasi-inverse $\hat{\phi} : X' \to X$ which is also a (C,D)-quasiisometry.

If x, y are two points in X which can be joined by a path $x = x_0, x_1, \dots, x_k = y$ outside of $B_v(n)$, then the images by ϕ of the vertices in this path lie outside of $B_{v'}(n/C - D)$. Now two vertices at distance one such as x_i and x_{i+1} can be joined by a path of length at most $C + D$ in X' - by joining these up we get a path between x and y which lies outside of $B_{v'}(n/C - D - (C + D))$. Thus every connected component of $X - B_v(n)$ is sent into a connected component of $X' - B_{v'}(f(n))$ where $f(n) = n/C - D - (C + D)$. It is easy to see that an unbounded component must be sent into an unbounded component.

Note that if $e(X) = e(X') = \infty$ we are done. We may thus assume that $e(X')$ is finite. In this case there exists $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$, we have $e_n(X') = e(X')$. Suppose wlog that $e(X) > e(X')$ - we can assume that for $n \geq n_0$ we have $e_n(X) > e(X')$.

Let us now see that for n such that $n, f(n) > n_0$, distinct unbounded components Y_1, Y_2 of $X - B_v n$ cannot be sent in the same connected component of $X' - B_{v'}(f(n))$. Pick two points $y_1 \in Y_1$ and $y_2 \in Y_2$ which cannot be joined by a path outside of $B_v(n)$. If $\phi(y_1)$ and $\phi(y_2)$ are in the same connected component of $B_{v'}(f(n))$, then they are in the same connected component of $B_{v'}(Cn+2CD)$ by assumption on n. Thus $\hat{\phi}(\phi(y_1))$ and $\hat{\phi}(\phi(y_2))$ can be joined by a path lying outside of $B_v(f(Cn + 2CD)) = B_v(n + D)$, which means that y_1 and y_2 can be joined by a path lying outside of $B_v(n)$ - a contradiction.

This proves the claim. $\square$

Thus in particular the number of ends of a Cayley graph is an invariant of the group - we thus define

Definition 2.42: *Let G be a finitely generated group. The number of ends of G is $e(G) = e(X(G,S))$ where S is a finite generating set for G*

Because of the symmetry inherent to Cayley graphs, a group cannot admit any number of ends.

Proposition 2.43: *(Freudenthal-Hopf theorem) Let G be a group. Then $e(G) \in \{0, 1, 2, \infty\}$.*

Proof. Denote by X a Cayley graph for G, and set v to be the vertex corresponding to the identity element. Suppose G has at least three ends, and let n be such that $e_n(X) \geq 3$ - the graph $X - B_v(n)$ contains at least 3 connected components X_0, X_1, X_2. Pick a vertex v_g corresponding to an element $g \in G$ in X_0 with $d(v, g) > 2n$. In the action of G on X, the element g sends the vertex v to v_g, in particular $X - B_{v_g}(n)$ has at least 3 unbounded connected components $g \cdot X_0, g \cdot X_1, g \cdot X_2$ - suppose wlog that $v \in g \cdot X_2$.

We now show that both X_0, X_1 are contained in $g \cdot X_0$. Indeed, any point in X_1, X_2 is joined to v by a path which does not pass through X_0, in particular it does not pass through $B_{v_g}(n)$.

Now set $N = d(v, v_g) + n$: we see that $X - B_v(N)$ contains at least 4 unbounded connected components - namely $X_1 - B_v(N), X_2 - B_v(N), g \cdot X_1 - B_v(N), g \cdot X_2 - B_v(N)$. This can be generalized to any number instead of 3. □

We can in fact characterize exactly what it means for a group to be in each one of these four possible cases.

Theorem 2.44: *(Stalling's theorem) Let G be a finitely generated group. Then*

1. *$e(G) = 0$ iff the group is finite;*
2. *$e(G) = 2$ iff the group is virtually $\mathbb{Z}$;*
3. *if $e(G) > 2$ then it splits as a non trivial amalgamated product over a finite group $G = A *_H B$.*

Thus a one-ended group is a group which is not virtually $\mathbb{Z}$ and does not admit a splitting as an amalgamated product over a finite group.

Note that if G is torsion free, the only finite subgroup is the trivial group.

2.4.2 Growth of a group

Definition 2.45: *Let G be a group endowed with a finite generating set S. The growth function of G relative to S is the function $\beta_{(G,S)} : \mathbb{N} \to \mathbb{N}$ defined by*

$$\beta_{(G,S)}(n) = \left|B_{(G,S)}(n) = \{g \in G \mid l_S(g) \leq n\}\right|$$

In other words, it is the cardinal of the ball of radius n in (G, l_S).

Example 2.46:

1. Growth of $F(S)$ with $|S| = k$: to choose a word of length l is choosing a reduced path in the Cayley graph of $F(S)$ with respect to S. Thus we have $(2k)(2k-1)\ldots(2k-1)$ possible choices (with $l-1$ ocurrences of $(2k-1)$). We get $\beta_{(F(S),S)}(n) = 1 + (2k)\sum_{l=1}^{n}(2k-1)^{l-1}$.
2. Growth of free abelian groups of rank 2 with canonical generating set: the ball of radius n in $\mathbb{Z}^2$ has cardinality $(n+1)^2 + n^2$ (can be decomposed into a square of side length $n+1$ and another one of side length n).

Exercise 2.47: Show that the growth function of the group $\mathbb{Z}^d$ relative to the canonical generating set is a polynomial of degree d.

Definition 2.48: *Let β_1, β_2 be two non decreasing functions $\mathbb{N} \to \mathbb{N}$. Say that β_2 dominates β_1 if there exists $A, B > 0$ such that $\beta_1(n) \leq A\beta_2(An + B) + B$ for all $n \in \mathbb{N}$.*

Say that β_1 and β_2 are equivalent if they dominate each other.

Note in particular that if β_1 is linear (resp. polynomial, exponential) then so is β_2.

Lemma 2.49: *If a group G embeds quasi isometrically in a group H then the growth function of H dominates the growth function of G.*

In particular, the equivalence class of the growth function is a quasiisometry invariant groups.

Proof. Let (G, l_S) and (H, l_T) be groups. Suppose $f : G \to H$ is a (C, D)-quasiisometric embedding.

First note that all the balls in a group endowed with the word metric have the same cardinality (left action by isometries).

Denote by B_n the ball of radius n around 1 in (G, d_S). Then $f(B)$ is contained in a ball of radius $Cn + D$ in (H, d_T), so $\beta_H(Cn + D) \geq |f(B_n)|$. On the other hand, if $f(g) = f(g')$ then $d_S(g, g') \leq C(d_T(f(g), f(g')) + D) = CD$ so $|B_n| \leq \beta_G(CD)\,|f(B_n)|$ so we get

$$\beta_G(n) = |B_n| \leq \beta_G(CD)\beta_H(Cn + D)$$

Since $\beta_G(CD)$ is a constant, we get our result. □

Remark 2.50: *In particular, if H has finite index in G then H and G have equivalent growth functions.*

Example 2.51: Fg infinite abelian groups have polynomial growth functions. Indeed, they contain with finite index a free abelian group.

Exercise 2.52: Show that the growth of finitely generated infinite nilpotent groups is bounded by a polynomial. [Recall that G is nilpotent if it admits a finite central series, i.e. a sequence

$$G = G_0 \geq G_1 \geq \ldots \geq G_r = 1$$

such that G_i/G_{i+1} is central in G/G_i for all i, i.e. for any elements $g \in G$ and $g_i \in G_i$, the commutator $[g, g_i]$ is in G_{i+1} (in particular, G_i/G_{i+1} is abelian).]

With some more work can show that a nilpotent group has polynomial growth. Gromov proved that the converse is (almost) true.

Theorem 2.53: *(Gromov's polynomial growth theorem) A finitely generated group has polynomial growth iff it is virtually nilpotent.*

3 Actions on trees

The aim of this section is to give an overview of Bass-Serre theory. The main result of Bass-Serre theory is that given an action of a group on a tree, one can give a presentation of G in terms of the vertex and edge stabilizers of the action.

3.1 Free groups and trees

We saw that the Cayley graph $X(F(S), S)$ of the free group $F(S)$ is a tree, so $F(S)$ acts on a tree. Moreover the action is free. Let us examine this action more closely.

What is its quotient? There is one orbit of vertices, and an orbit of edge for each element in S. Thus the quotient is a rose graph: one vertex, and $|S|$ loops.

Note that the quotient map is a covering map (look at the neighborhood of a point in the interior of an edge, and at a neighborhood of the vertex).

This is an instance of a more general fact

Proposition 3.1: *Let X be a Hausdorff and path connected space. Let G be a group which acts freely and properly discontinuously on X. Then $p : X \to G\backslash X$ is a covering map.*

Remark 3.2: *In fact, we have seen that if a group G acts on a graph with finite vertex stabilizers then the action properly discontinuous, so if X is a graph freeness automatically implies proper discontinuity.*

Let us go back to our specific example of $F(S)$ acting on its Cayley graph. Since $X(F, S)$ is a tree, it is simply connected so this is in fact a universal cover.

The following is a fundamental theorem of covering space theory.

Theorem 3.3: *Let X be a nice (Hausdorff, connected, locally path connected) simply connected space, let $x \in X$. If a group G acts on X in a free and properly discontinuous way, G is isomorphic to $\pi_1(X\backslash G, p(x))$ via the application which to $g \in G$ associates $p \circ \gamma_g$ where γ_g is a path joining x to $g \cdot x$.*

The theorem thus implies that $F(X, S)$ is the fundamental group of the rose, and that the loops corresponding to the elements of S are exactly the petals of the rose.

Remark 3.4: *Let Y be a graph. By collapsing a maximal subtree Y_0 of Y, one gets a space which is homotopy equivalent to a rose, hence which has a free fundamental group. Thus the fundamental group of a graph is always free, of rank corresponding to the number of edges that lie outside of a maximal subtree.*

Theorem 3.5: *A group acts freely on a tree iff it is free.*

Proof. Since we have seen that the free group $F(S)$ acts freely on its Cayley graph with respect to S, which is a tree, there remains only one direction to prove. By Remark 3.2, if G acts freely on the tree Y then the quotient map $Y \to G\backslash Y$ is a covering map. Since Y is simply connected, this is in fact a universal cover so we have $\pi_1(G\backslash Y) = G$. But $G\backslash Y$ is a graph, hence its fundamental group is free. □

Note that is also possible to give an elementary proof which does not rely on covering space theory.

Corollary 3.6: *Any subgroup of a free group is free.*

Proof. Let H be a subgroup of a free group G. The group G acts freely on its Cayley graph X which is a tree. Thus H also acts on X, and the action is also free (if points don't have stabilizers in the big group they don't have any in the small group). Hence H is itself free. □

A useful point of view on free groups. (Unlectured) Seeing the free group as the fundamental group of the rose enables one to use the power of covering space theory to deduce algebraic properties of the free groups - we give an example.

Recall that if $p : X \to Y$ is a covering map between path connected spaces, then it induces an injective map $p_* : \pi_1(X, x) \to \pi_1(Y, p(x))$ given by $p_*([\gamma]) = [p \circ \gamma]$. Moreover if p is an n-sheeted cover then $p_*(\pi_1(X, x))$ has index n in $\pi_1(Y, p(x))$.

Example 3.7: Consider a finite sheeted cover $q : \hat{R} \to R$ of the rose with 2 petals. Its fundamental group is a finite index subgroup of the free group $F(a, b)$. But its fundamental group is itself free, since $\hat{R}$ is a graph. One can see that an n sheeted cover will have fundamental group free of rank $n+1$ (by counting vertices and edges - know the number of edges not in a maximal subtree).

For example, draw the cover corresponding to subgroup generated by

$$\{a^n, b, aba^{-1}, a^2ba^{-2}, \dots, a^{n-1}ba^{1-n}\}.$$

Thus for any $m \geq 2$, the free group $\mathbb{F}_m$ embeds in $\mathbb{F}_2$ as a subgroup of finite index.

3.2 Non free actions on trees

What happens if G acts on a tree, but not freely?

Amalgamated product case

Proposition 3.8: *Suppose G acts on a tree T without inversion so that there is one orbit of edges and two orbits of vertices. Then G can be written as an amalgamated product* $\mathrm{Stab}(p) *_{\mathrm{Stab}(e)} \mathrm{Stab}(q)$ *where e is an edge of T with endpoint p, q.*

We gave a constructive definition of amalgamated products:

Definition 3.9: *Let $A = \langle S_A \mid R_A \rangle$, $B = \langle S_B \mid R_B \rangle$, C a group with embeddings $C \to A, c \mapsto c_A$, $C \to B, c \mapsto c_B$. The amalgamated product of A with B over C is the quotient of $A * B$ by the normal subgroup generated by the elements $c_A c_B^{-1}$ for all $c \in C$. It is denoted by $A *_C B$ (abuse of notation).*

To prove that a given group is an amalgamated product of two of its subgroups, we will use the following lemma:

Lemma 3.10: *Let G be a group, let A, B be subgroups of G and $C = A \cap B$. If*

1. *A and B generate G;*
2. *no alternating product of elements of $A - C$ and $B - C$ (i.e. for example of the form $a_1 b_1 \dots b_k a_k$ with $a_i \in A - C, b_i \in B - C$) is trivial.*

then G is the amalgamated product of A and B over C.

Proof. By definition of amalgamated products, there is a surjective morphism $p : A * B \to A *_C B$ whose kernel is the normal closure of $R = \{c_A c_B^{-1} \mid c \in C\}$.

On the other hand, by the universal property of free products, there is a unique morphism $\pi : A * B \to G$ induced by the inclusions of A, B in G. Condition 1. ensures that π is surjective. Clearly $R \subseteq \mathrm{Ker}\pi$, so π factors through p as $\pi = \pi' \circ p$. We now want to show that π' is in fact injective.

Note that for any $w \in A * B$, there exists $w' = (a_0) b_1 a_1 \dots b_k a_k (b_{k+1})$ with $a_i \in A - C$ and $b_i \in B - C$ such that w and w' represent the same element of $A *_C B$, so they have the same image by p and thus also by π - in fact, this is not quite true because w could be just an element of C. In summary w' is either an alternating product of elements of $A - C$ and $B - C$, or an element of C itself.

Now if $w \in \mathrm{Ker}\pi$, so is w'. By condition 2., we must therefore have that w' is an element of C. But the only such element in $\mathrm{Ker}\pi$ is 1. Thus $p(w) = p(1) = 1$. Thus we proved $\mathrm{Ker}\pi = \mathrm{Ker}p$ - this implies $p = \pi$ hence $G = A *_C B$. □

We first prove the following lemma.

Lemma 3.11: *Show that $g \cdot p$ is at distance 1 of $h \cdot q$ if and only if $g = hvu$ where $v \in \mathrm{Stab}(q)$ and $u \in \mathrm{Stab}(p)$ if and only if $h = gu'v'$ where $v' \in \mathrm{Stab}(q)$ and $u' \in \mathrm{Stab}(p)$.*

Proof. The second equivalence follows by taking $v' = v^{-1}$ and $u' = u^{-1}$.

It is easy to see that $vu \cdot p$ is at distance 1 of q. Hence $hvu \cdot p$ is at distance 1 of $h \cdot q$.

To prove the converse, we first translate by h^{-1}: note that there is an edge f between $h^{-1}g \cdot p$ and q. Since there is only one orbit of edges, there is an element sending e to f: this element v must fix q, so $v \in \mathrm{Stab}(q)$, and send p to $h^{-1}g \cdot p$. Hence $h^{-1}g \cdot p = v \cdot p$, thus $v^{-1}h^{-1}g = u \in \mathrm{Stab}(p)$: this proves the result. $\square$

We can now prove the Proposition.

Proof. We first show that Stab(p) and Stab(q) generate. Let $g \in G$. Consider the path $p, h_1 \cdot q, g_1 \cdot p, h_2 \cdot q, \ldots, h_k \cdot q, g_k \cdot p = g \cdot p$ between p and $g \cdot p$. By the lemma above, $g = h_k vu$ for some v in Stab(q) and $u \in \mathrm{Stab}(p)$, hence by induction the result is proved.

Consider an alternating product $u_0 v_1 u_1 \ldots v_k u_k v_{k+1}$ where $u_i \in \mathrm{Stab}(p)$, $v_i \in \mathrm{Stab}(q)$, and $u_i, v_i \notin \mathrm{Stab}(e)$ for any $1 \leq i \leq k$.

The sequence $p, u_0 \cdot q, u_0 v_1 \cdot p, \ldots, u_0 v_1 \ldots v_k u_k v_{k+1} \cdot p = p$ defines a path in X by the Lemma above. If $u_0 v_1 u_1 \ldots v_k u_k v_{k+1}$ represents the trivial element in G, in fact it is a cycle. Now T is a tree, so the path must backtrack at some point: wlog we have $u_0 v_1 u_1 \ldots u_{i-1} v_{i-1} \cdot p = u_0 v_1 u_1 \ldots u_i v_i \cdot p$. Hence we get $p = u_i v_i \cdot p$, that is, $u_i v_i = u$ for some $u \in \mathrm{Stab}(p)$. Thus $u^{-1} u_i = v_i$ which contradicts the hypotheses on the u_i's.

Hence no such alternating product represents the trivial element, and G is the amalgamated product $\mathrm{Stab}(p) *_{\mathrm{Stab}(e)} \mathrm{Stab}(q)$. $\square$

In fact, the converse is true:

Proposition 3.12: *If $G = A *_C B$, then G admits an action on a tree whose quotient is a segment, in which there is an edge $e = (p, q)$ with* Stab(e) $= C$, Stab(p) $= A$ *and* Stab(q) $= B$.

HNN extension case We have a vsimilar result corresponding to HNN extensions.

Proposition 3.13: *If a group G acts on a tree T without inversion so that there is one orbit of edges and one orbit of vertices, then G can be written as an HNN extension* $\mathrm{Stab}(p)*_{\mathrm{Stab}(e)}$ *with stable letter t, where e is an edge of T with endpoint $p, t \cdot p$.*

Also here the converse holds

Proposition 3.14: *If $G = A*_C$, then G admits an action on a tree whose quotient is a loop graph (one edge and one vertex), such that the vertex stabilizers are conjugate to A, and the edge stabilizers are conjugate to C.*

Remark 3.15: *Suppose G acts on a tree T without inversion, and pick an edge e of T: by collapsing all the edges which are not in the orbit of e, we get a new tree T' on which G still acts, and the stabilizer of the image e' of e in T' is still* Stab(e). *By the two propositions above, G admits a splitting (either as an amalgamated product or as an HNN) over* Stab(e).

Hence if G acts on a tree T without inversion, it admits splittings over each of the edge stabilizers.

Remark 3.16: *Special case: if G acts on a tree with trivial edge stabilizers, then G splits as a free product.*

More generally, given an action of a group G without inversions on a tree T, Bass-Serre theory gives a presentation of G in terms of the stabilizers of edges and vertices of T.

4 Hyperbolic groups

Around 300BC, the hellenistic mathematician Euclid wrote his (world famous) book "The elements", which contains the first systematic treatement of geometry.

In Book I, he gives a number of definitions for points, lines, etc. and five axioms:

1. Each pair of points can be joined by one and only one straight line segment.
2. Any straight line segment can be indefinitely extended in either direction.
3. There is exactly one circle of any given radius with any given center.
4. All right angles are congruent to one another.

and a fifth which is equivalent (given the other four) to

5' Given a line and a point not on it, there is exactly one line going through the given point that is parallel to the given line.

This last axiom, often called "the parallel axiom", looks much more complicated than the other four, and mathematicians tried to deduce it from the other axioms for centuries, mostly by replacing it by other axioms which they thought were simpler. In the 19th century, people (Gauss, Bolyai, Lobachevsky) began to try to prove this by contradiction: if we assume that given a point p and a line L, there are infinitely many line going through p which are parallel to (i.e. do not meet) L, what happens? But instead of running into a contradiction, these mathematician worked out more and more results of this alternate geometry, and began to become convinced of its consistency.

The consequences were strange to the Euclidean mind: one would get that the set of points at a fixed distance of a line was not itself a line, that the sum of the angles of a triangle was always strictly less than π (in fact, the fact that the sum of the angles of a triangle is equal to π is equivalent to the parallel axiom).

The study of curved surfaces (ex: the two-holed torus) laid the basis for construction of analytic models of this new set of axioms. Other models, including the one we give below, were worked out by Henri Poincare at the end of the 19th century/beginning of the 20th. A central such model is that of the half hyperbola, which gave its name to this particular flavor of non-euclidean geometry: hyperbolic geometry.

Hyperbolic geometry is central in the works of Minkowski and Einstein on relativity. For more on the historical development of hyperbolic geometry see the first section of "Hyperbolic geometry" by Cannon, Floyd, Kenyon and Parry (available online).

But before we give an explicit model of hyperbolic geometry, here is another model of non-euclidean geometry that had in fact be known for centuries before...(note though that it does not satisfy the first four axioms of Euclid so its status is a bit different)

Example 4.1: (Spherical geometry) Let $\mathbb{S}^2$ denote the sphere in $\mathbb{R}^3$. We call lines on $\mathbb{S}^2$ the great circles - they are the curves minimizing distance locally, so they are indeed the analogues of lines in Euclidean plane in that sense.

In this space, given a point p and a line L, how many lines going through p are parallel (i.e. non-intersecting) to L? None! Any two great circles intersect.

Note also that the sum of angles of a triangle on a sphere is always greater than π (consider for example a triangle with a right angle at the North pole and two points on the equator: the sum of its angle is $3\pi/2$).

The following metric space is a model of "hyperbolic geometry", that is, it is an example of a space which satisfies the first four axioms of Euclid, but in which given a point and a line there are infinitely many lines going through the point which do not meet the given line.

Example 4.2: Let $\mathbb{H}^2$ denote the upper half plane, that is, the set of points $\{(x,y) \in \mathbb{R}^2 \mid y > 0\}$. We want to endow $\mathbb{H}^2$ with a metric, which we will call the hyperbolic metric.

We first define the hyperbolic length of a curve $\gamma : [a, b] \to \mathbb{H}^2$ (i.e. a $\mathcal{C}^1$ map) by the following formula:

$$l_h(\gamma) = \int_a^b \frac{|\gamma'(t)|}{\gamma_y(t)} dt$$

We can then define a function

$$d_h(p, q) = \inf\{l_h(\gamma) \mid \gamma \text{ a piecewise } \mathcal{C}^{\ 1} \text{ curve joining } p \text{ to } q \text{ in } \mathbb{H}^2\}$$

It is possible to show that this defines a metric on $\mathbb{H}^2$.

It can also be shown that the shortest curve joining two points is either a piece of vertical line, or a piece of a half-circle centered on the real axis - if we take these to be the "lines" of our space, the first four axioma of Euclid are indeed satisfied. Moreover, it is easy to see that there are infinitely many parallels to a given line through a given point.

Another model of the same space is the Poincare disk:

$$\mathbb{D}^2 = \{z \in \mathbb{C} \mid |z| = 1\}$$

endowed with the metric $d_{\mathbb{D}^2}$ induced by the length function for curves given by

$$l_h(\gamma) = \int_a^b \frac{|\gamma'(t)|}{1 - (\gamma_x^2(t) + \gamma_y^2(t))} dt$$

The two models are isometric via the map $(\mathbb{H}^2, d_{\mathbb{H}^2}) \to (\mathbb{D}^2, d_{\mathbb{D}^2})$ defined by $z \mapsto \frac{i-z}{zi-1}$. Lines are arcs of circles orthogonal to the boundary of the disks as well as diameters.

See Escher representations of these two models.

4.1 Hyperbolic metric spaces

One can also try to build combinatorial models that capture the behaviour of hyperbolic geometry, which we can think of as approximation to hyperbolic geometry. One such attempt was made by Gromov in the 80's, who gave a definition of hyperbolicity for a metric space. The definition we give is not Gromov's original one, but is equivalent to it when the metric space is geodesic. It is attributed to Rips.

Recall Definition 2.17. Let X be a geodesic metric space.

Definition 4.3: *A geodesic triangle in X is a triple $(\gamma_0, \gamma_1, \gamma_2)$ of geodesic segments $\gamma_i : [0, L_i] \to X$ such that $\gamma_0(L_0) = \gamma_1(0)$, $\gamma_1(L_1) = \gamma_2(0)$, $\gamma_2(L_2) = \gamma_0(0)$.*

Definition 4.4: *Let A be a subset of a metric space X, and let $\epsilon \geq 0$. The ϵ-neighborhood of A in X is given by*

$$B_\epsilon(A) = \{x \in X \mid \exists a \in A \text{ such that } d(x, a) < \epsilon\}.$$

Definition 4.5: *Let $\delta \geq 0$. A geodesic triangle $(\gamma_0, \gamma_1, \gamma_2)$ is said to be δ-slim if each of its sides is contained in a δ-neighbourhood of the union of the two other sides.*

The space X is said to be δ-hyperbolic if all its geodesic triangles are δ-slim.

Example 4.6: Any metric space with finite diameter D is D-hyperbolic.

Example 4.7: $\mathbb{R}^2$ with its usual metric is not hyperbolic: consider the triangle between $(0, 0)$, $(0, 3\delta)$ and $(3\delta, 0)$.

Example 4.8: A metric tree (built from a graph-tree by identifying every edge with the interval $[0, 1]$) is 0-hyperbolic (triangles are tripods).

In fact, the converse is almost true: if X is a geodesic 0-hyperbolic space, then X is a real tree.

Definition 4.9: *An arc in a metric space X joining points p and q is the image of a continuous injective map $\gamma : [0, a] \to X$ so that $\gamma(0) = p$ and $\gamma(a) = q$.*

A real tree is a geodesic metric space in which any two points are joined by a unique arc.

Example 4.10: • SNCF metric on $\mathbb{R}^2$: $d(u,v) = |v-u|$ if $v = \lambda u$, and $d(u,v) = |u| + |v|$ if not.

- Comb metric on $\mathbb{R}^2$: $d(u,v) = |y_v - y_u|$ if $x_v = x_u$, and $d(u,v) = |y_u| + |x_v - x_u| + |y_v|$ if not.

How about a non bounded example where $\delta \neq 0$?

Lemma 4.11: *The hyperbolic plane $\mathbb{H}^2$ is δ-hyperbolic, and the best hyperbolicity constant δ is* $\ln(1+\sqrt{2})$.

Want to show that the notion of hyperbolicity is invariant by quasi-isometry but for this need notion of quasigeodesic. Problem: the image by a quasi-isometry of a geodesic segment is not necessarily a geodesic.

Definition 4.12: *A (C,D)-quasi-isometric embedding $c : I \to X$ is called*

- *a (C,D)-quasigeodesic segment in X if $I = [a,b]$;*
- *a (C,D)-quasigeodesic ray if $I = [0,\infty)$;*
- *a (C,D)-quasigeodesic (or quasigeodesic line) if $I = \mathbb{R}$.*

Remember the logarithmic spiral (section on quasi-isometries): it is a quasigeodesic.

Exercise 4.13: Geodesic segments are exactly $(1,0)$-quasigeodesics.The image of a geodesic segment/ray/line by a (C,D)-quasi-isometry is a (C,D)-quasigeodesic. The image of a quasigeodesic segment/ray/line by a quasi-isometry is a quasigeodesic segment/ray/line.

Theorem 4.14: *(Stability of quasigeodesic segments). Let X be a δ hyperbolic space. Let $C \geq 1, D \geq 0$. There exists a constant $R = R(C,D,\delta)$ such that for any (C,D)-quasigeodesic segment $\gamma : [a,b] \to X$, if c is a geodesic segment joining $\gamma(a)$ to $\gamma(b)$, then $\mathrm{Im}(\gamma)$ remains within a R-neighborhood of $\mathrm{Im}(c)$, and $\mathrm{Im}(c)$ remains within a R-neighborhood of $\mathrm{Im}(\gamma)$.*

To prove this we need the following lemma which enables us to replace a quasigeodesic by a "nicer" one:

Lemma 4.15: *Let X be a geodesic space. Let $\gamma : [a,b] \to X$ be a (C,D)-quasigeodesic segment. There exists a* **continuous** *map $\gamma_1 : [a,b] \to X$ such that*

(i) *$\gamma(a) = \gamma_1(a)$ and $\gamma(b) = \gamma_1(b)$;*

(ii) *γ_1 is a (C,D')-quasigeodesic segment with $D' = 2C + 3D$;*

(iii) *$l(\gamma_1 |_{[s,t]}) \leq K_1 d(\gamma_1(s), \gamma_1(t)) + K_2$ for any $s,t \in [a,b]$ and constants K_i which depend only on C and D;*

(iv) *the images of γ and γ_1 are contained in $(C+D)$-neighborhoods of one another.*

What do we mean in (iii) when we talk about the length of $\gamma_1 |_{[s,t]}$?

Definition 4.16: *Let $\delta : [a,b] \to X$ be a continuous map. We define the length of δ to be*

$$l(\delta) = \sup\{\sum_{i=1}^{n} d_X(\delta(a_{i-1}), \delta(a_i)) \mid a = a_0 < a_1 < \ldots < a_n = b, n \in \mathbb{N}\}$$

If this sup is not infinite, say that δ is rectifiable.

Proof of the lemma:

Proof. We define γ' by a concatenation of geodesic segments joining the points $\gamma(a), \gamma(m), \gamma(m+1), \ldots, \gamma(m+p), \gamma(b)$ where $\{m, \ldots, m+p\}$ are the integer points of $[a,b]$. We parameterize it continuously by a map $\gamma_1 : [a,b] \to X$ such that for each $k \in \{a, m, \ldots, m+p, b\}$ we have $\gamma(k) = \gamma_1(k)$. In particular, (i) is satisfied.

For any $t \in [a,b]$, let $[t]$ denote a point of $a, m, m+1, \ldots, m+p, b$ closest to t. For any t in $[a,b]$ we have $|t - [t]| \leq 1/2$, hence any point $\gamma(t)$ on $\gamma([a,b])$ is at a distance at most $C/2 + D$ of $\gamma([t])$, which lies on $\mathrm{Im}(\gamma_1)$. On the other hand, for each $m \leq k \leq m+p$, we have $d(\gamma(k), \gamma(k+1)) \leq C + D$ since γ is a (C,D)-quasigeodesic. Hence any point on $\gamma_1([a,b])$ is at a distance at most $(C+D)/2$ of a point of the form $\gamma(k)$ or $\gamma(a), \gamma(b)$. Thus we see that (iv) is satisfied.

Note that $\gamma_1([t]) = \gamma([t])$ for all t. We get

$$\begin{array}{rcl} d(\gamma_1(t), \gamma_1(s)) & \leq & d(\gamma_1([t]), \gamma_1([s])) + C + D \\ & = & d(\gamma([t]), \gamma([s])) + C + D \\ & \leq & C\,|[t] - [s]| + D + (C+D) \\ & \leq & C(|t-s| + 1) + C + 2D \\ & \leq & |t-s| + 2C + 2D \end{array}$$

and

$$\begin{array}{rcl} (1/C)\,|t-s| - D & \leq & d(\gamma(t), \gamma(s)) \\ & \leq & d(\gamma([t]), \gamma([s])) + C + 2D \\ & = & d(\gamma_1([t]), \gamma_1([s])) + C + 2D \\ & \leq & d(\gamma_1(t), \gamma_1(s)) + C + 2D + C + D \end{array}$$

so γ_1 is a $(C, 2C+3D)$-quasigeodesic segment which proves (ii).

To prove (iii), first note that for every integers k, l in $[a,b]$ we have:

$$l(\gamma_1\,|_{[k,l]}) = \sum_{j=k}^{l-1} d(\gamma_1(j), \gamma_1(j+1)) \leq |l-k|\,(C+D)$$

and $l(\gamma_1\,|_{[a,l]}) \leq (l-a+1)(C+D)$ and $l(\gamma_1\,|_{[k,b]}) \leq (b-k+1)(C+D)$. So for any t, s in $[a,b]$ we have

$$l(\gamma_1\,|_{[t,s]}) \leq l(\gamma_1\,|_{[[t],[s]]}) + (C+D) \leq (|[t]-[s]| + 1)(C+D)$$

On the other hand we proved that

$$d(\gamma_1(t), \gamma_1(s)) \geq (1/C)\,|t-s| - (2C+4D) \geq (1/C)(|[t]-[s]| - 1) - (2C+4D)$$

So $|[t]-[s]| \leq C d(\gamma_1(t), \gamma_1(s)) + C(2C+4D) + 1$ and finally we get

$$l(\gamma_1\,|_{[t,s]}) \leq C(C+D) d(\gamma_1(t), \gamma_1(s)) + C(C+D)(2C+4D) + 2(C+D).$$

□

We need another lemma, which shows that geodesics stay close to short curves.

Lemma 4.17: *Suppose X is a hyperbolic metric space. Let $\gamma : [a,b]$ be a continuous path with $l(\gamma) < \infty$ (rectifiable). For any x on a geodesic segment $[\gamma(a), \gamma(b)]$, we have*

$$d(x, \mathrm{Im}(\gamma)) \leq \delta\,|\log_2(l(\gamma))| + 1$$

Note that this does not hold in Euclidean space: consider a right angle triangle ABC of sides AB and BC of lengths x and AC of length $\sqrt{2x^2}$: the midpoint of AC is at distance $\sqrt{2}x/2$ of any other point.

Proof. Assume without loss of generality that γ is parametrized proportionally to length (i.e. that there exists α with $l(\gamma\mid_{[s,t]}) = \alpha\,|t-s|$ for all s,t).

Pick $N \in \mathbb{N}$ such that $l(\gamma) \leq 2^N$. We show by induction on N that $l(\gamma) \leq \delta N + 1$.

For $N = 0$, it is clear (then $d(\gamma(a), \gamma(b)) \leq 1$). Suppose it is true for N, and let x be a point on $[\gamma(a), \gamma(b)]$, and consider the triangle $\gamma(a), \gamma(a/2 + b/2), \gamma(b)$. Some point x' on (wlog) the side $[\gamma(a), \gamma(a/2 + b/2)]$ is at a distance at most δ from x. By induction, x' is at a distance at most $(N-1)\delta + 1$ from $\mathrm{Im}(\gamma)$. This proves the claim. □

We can now prove Theorem 4.14.

Proof. We have a (C,D)-quasigeodesic segment $\gamma : [a,b] \to X$, and a geodesic segment c joining $\gamma(a)$ to $\gamma(b)$. By Lemma 4.15 above we can assume that γ is continuous and satisfies conditions $(i) - (iv)$.

Let $M_\gamma = \sup\{d(x, \mathrm{Im}(\gamma)) \mid x \in \mathrm{Im}(c)\}$, and let x_0 be the point at which this sup is reached. Let y and z be the points of the chosen geodesic segment $[\gamma(a), \gamma(b)]$ at a distance exactly $2M_\gamma$ from x_0 (or $\gamma(a), \gamma(b)$ respectively if their distance from x_0 is less than $2M_\gamma$). Let y', z' be points on $\mathrm{Im}(\gamma)$ at distance $\leq M_\gamma$ from y, z.

Consider the path β from y to z going through a geodesic segment between y and y', then the path γ up to z', and back to z through a geodesic segment between z' and z. By part (3) of the Lemma above we get that

$$l(\beta) \leq l(\gamma\mid_{[y',z']}) + 2M_\gamma \leq K_1(6M_\gamma) + K_2 + 2M_\gamma$$

By the Lemma 4.17 we get:

$$M_\gamma \leq \delta\,|\log_2(l(\beta))| + 1 \leq \delta\,|\log_2(K_1(6M_\gamma) + K_2 + 2M_\gamma)| + 1$$

The RHS grows logarithmically with M_γ while the LHS grows linearly, hence M_γ must be bounded above (for all curves γ) by some constant M_0 which depends only on δ, C, and D. Thus $[\gamma(a), \gamma(b)]$ lies in $B_{M_0}(\mathrm{Im}(\gamma))$.

Let now u be such that $d(\gamma(u), \mathrm{Im}(c))$ is maximal. Now $\mathrm{Im}(c)$ is contained in $B_{M_0}(\gamma([a,u))) \cup B_{M_0}(\gamma((u,b]))$. By connectedness of $\mathrm{Im}(c)$ there must be a point w of $\mathrm{Im}(c)$ which lies in both neighborhoods, that is, at distance $\leq M_0$ of both $\gamma(s)$ and $\gamma(t)$ for $s < u < t$. The distance $d(\gamma(s), \gamma(t))$ is at most $2M_0$, hence the length of $\gamma\mid_{[t,s]}$ is at most $K_1 2M_0 + K_2$. Hence any point of $\gamma\mid_{[t,s]}$, in particular $\gamma(u)$ is at a distance at most $K_1 M_0 + K_2/2 + M_0$ of w. Thus $\mathrm{Im}(\gamma)$ is contained in the $(K_1 M_0 + K_2/2 + M_0)$-neighborhood of $[\gamma(a), \gamma(b)]$.

□

This enables us to prove the following corollaries.

Corollary 4.18: *A metric space Y is hyperbolic if and only if for any constants $C \geq 1, D \geq 0$ there exists $M(C,D)$ such that any (C,D)-quasigeodesic triangle is M-thin.*

Proof. One direction is obvious since geodesic triangles are $(1,0)$-quasigeodesic triangles. Let $R = R(\delta, C, D)$ be such that any (C,D)-quasigeodesic segment in Y is within an R-neighborhood of any geodesic segment joining its endpoints. Let $\gamma_0, \gamma_1, \gamma_2$ be a (C,D)-quasigeodesic triangle. Let $\beta_0, \beta_1, \beta_2$ be geodesic segments joining the endpoints of $\gamma_0, \gamma_1, \gamma_2$. A point x on γ_0 is at distance at most R from a point on β_0, which is itself at distance at most δ from a point of β_1 (wlog), which in turn lies within distance R of a point y of γ_1. Setting $M = 2R + \delta$ proves the claim. □

Corollary 4.19: *Suppose that $f : X \to Y$ is a (C,D)-quasi-isometric embedding. If Y is δ-hyperbolic, then X is δ'-hyperbolic for some δ' which depends only on δ, C, D.*

Proof. Let $\gamma_0, \gamma_1, \gamma_2$ be a geodesic triangle in X. The (C,D)-quasigeodesic triangle $f \circ \gamma_0, f \circ \gamma_1, f \circ \gamma_2$ is M-thin by the previous lemma so any point $f(x)$ on $f \circ \gamma_0$ is at distance at most M of $f(y)$ for some point y on γ_1 or γ_2. Since f is a (C,D)-quasiisometric embedding, we get $d(x,y) \leq Cd(f(x), f(y)) + D \leq CM + D$. Setting $\delta' = CM + D$ finishes the proof. □

4.2 Hyperbolic groups - definition

The following definition is due to Rips and Gromov.

Definition 4.20: *We say that a group G is hyperbolic if its Cayley graph (endowed with the metric which identifies each edge to the interval $[0,1]$) is a hyperbolic metric space.*

By the theorem above, this does not depend on the choice of a generating set.

Example 4.21:
1. Every finite group is hyperbolic.
2. Free groups are hyperbolic.
3. $\mathbb{Z}^2$ is NOT hyperbolic.
4. $\mathrm{SL}_2(\mathbb{Z})$ is hyperbolic.
5. The fundamental groups of orientable surfaces of genus at least 2 is hyperbolic (such surfaces admit metrics for which the universal cover is isometric to H^2 - but the fundamental group, as we saw, acts properly discontinuously cocompactly on the universal cover in this case, hence it is quasi-isometric to $\mathbb{H}^2$).

4.3 Finite presentation of hyperbolic groups

The following proposition shows that in a sense, hyperbolicity is a strong constraint on a group: remember that we saw that there are only countably many finitely presented groups, while the class of finitely generated groups is uncountable.

Proposition 4.22: *Hyperbolic groups are finitely presented.*

We need the following lemma, which shows that geodesic segments with close endpoints remain close throughout. Note that this is more precise than Theorem 4.14, since we even know which point of c' is close to $c(t)$.

Lemma 4.23: *Let $c : [0,T] \to X$ and $c' : [0,T'] \to X$ be geodesic segments. Suppose that $c(0) = c'(0)$. Then for any $t \geq 0$ we have that*

$$d(c(t), c'(t)) \leq 2(\delta + d(c(T), c'(T')))$$

where for any $t > T$, we define $c(t)$ to be equal to $c(T)$.

Proof. Let $t' \in [0,T']$ and let $D = d(c(t), c'(t'))$. Then the triangle inequality on $c(0) = c'(0), c(t), c'(t')$ gives

$$|d(c(0), c(t)) - d(c'(0), c'(t'))| = |t' - t| \leq d(c(t), c'(t'))$$

Hence $d(c(t), c'(t)) \leq d(c(t), c'(t')) + d(c'(t'), c'(t)) \leq 2d(c(t), c'(t'))$.

Consider "the" geodesic triangle between $c(0), c(T)$ and $c'(T')$, and a point $c(t)$. If $c(t)$ is δ-close to some point $c'(t')$ on the side $[c(0), c'(T')]$, then $d(c(t), c'(t)) < 2\delta$. If not, then $c(t)$ is within distance δ of the side between $c(T)$ and $c'(T')$. Thus $c(t)$ is within $\delta + d(c(T), c(T'))$ of $c'(T')$. We get $d(c(t), c'(t)) < 2(\delta + d(c(T), c(T')))$. □

We prove finite presentation of hyperbolic groups:

Proof. Let S be a finite generating set for G. All the relations between the elements of S can be written as products of relations of the form csd^{-1} where s is in S and c, d are geodesic words (a word $s_1 \dots s_r$ is said to be geodesic if the path given by following the edges labelled $s_1, \dots, s_r$ in the Cayley graph of G with respect to S is geodesic).

Hence it suffices to show that finitely many relations imply all the relations of this type.

Let w be a relation between the elements of S of this form, $w = csd^{-1}$, with $c = u_1 \dots u_l$ and $d = v_1 \dots v_m$ for $u_i, v_j \in S \cup S^{-1}$. Note that $|l - m| \leq 1$ (triangle inequality), wlog assume that $l \leq m$.

Let α_n be a geodesic word such that $\alpha_n =_G c_n^{-1} d_n$. By the previous Lemma we know that the length of α_n is at most $2\delta + 2$.

Note that

$$c_n \alpha_n d_n^{-1} = c_{n-1} u_n \alpha_n v_n^{-1} d_n^{-1} = [c_{n-1} \alpha_{n-1} d_{n-1}^{-1}][d_{n-1}(\alpha_{n-1} u_n \alpha_n v_n^{-1}) d_{n-1}^{-1}]$$

and that $\alpha_{n-1} u_n \alpha_n v_n^{-1}$ is a relation in G of length at most $4\delta + 6$.

By induction on n, we can show that $c_l \alpha_l d_l^{-1}$ can be written as a product of conjugates of relations of length at most $4\delta + 6$. If $m = l$ we are done since $c_l = c$, wlog $\alpha_l = s$ and $d_l = d_m = d$. If $m = l + 1$ we have

$$csd^{-1} = (c_l \alpha_l d_l^{-1}) d_l (\alpha_l^{-1} s v_{l+1}^{-1}) d_l^{-1}$$

and $(\alpha_l^{-1} s v_{l+1}^{-1})$ is a relation of length at most $2\delta + 4$ in G.

Thus the relations of lengths at most $4\delta + 6$ imply all the other relations of G: since there are finitely many such relations, we have found a finite presentation for G. □

Note that in fact we proved something stronger: starting from any finite set of generators for G, we can find a finite presentation with this set of relators.

Though it may seem hard to come by examples of hyperbolic groups, among finitely presented groups, hyperbolic groups are everywhere:

Theorem 4.24: *A "random" finitely presented group is hyperbolic.*

This can be shown for several randomness notions, for example: one fixes the number n of generators and m of relators, and on considers all finite presentations of groups with n generators and m relations of length at most t on these generators. If we denote by $P(t)$ the probability that a presentation chosen uniformly among these gives a hyperbolic group, then $P(t) \to 1$ as $t \to \infty$.

In fact, a famous statement of Gromov is that any property which is true of all groups is trivial, and this is what prompted him to look for a class of groups which would be very generic, yet still interesting.

4.4 Word problem in hyperbolic groups

This puts in perspective the following result of Gromov, which we will now prove:

Theorem 4.25: *Hyperbolic groups have solvable word problem.*

In fact, what we show is a weak version of this: every hyperbolic group admits a presentation for which the word problem is solvable.

The proof relies on an idea of Dehn.

Definition 4.26: *Let S be an alphabet. A reduced word w in S is cyclically reduced if $w = s_1 s_2 \dots s_n$ with $s_1 \neq s_n^{-1}$.*

Note that any reduced word admits a cyclically reduced conjugate, and that this can be computed algorithmically.

Definition 4.27: *A presentation $\langle S \mid R \rangle$ is said to be symmetric if for any $r \in R$, r is cyclically reduced and all the cyclically reduced conjugates of r, r^{-1} are in R.*

Exercise 4.28: Given a finite presentation $\langle S \mid R \rangle$, describe an algorithm which computes a finite presentation $\langle S \mid R' \rangle$ of the same group which is moreover symmetric.

Definition 4.29: *A finite symmetric presentation $\langle S \mid R \rangle$ of a group is called a Dehn presentation if any word w on S which represents the identity in G contains a subword v which is more than half of a relator, more precisely, for some r in R the word v is also a subword of r and we have $r = uv$ with u a subword of r satisfying $l_S(u) < l_S(v)$.*

Note that the symmetry condition is not always included, it just makes the proofs neater.

Proposition 4.30: *(Dehn's algorithm). If $\langle S \mid R\rangle$ is a Dehn presentation then the word problem for the group $G = \langle S \mid R\rangle$ is solvable.*

Proof. Consider the following algorithm: First, make a list L of all the pairs (v, u) such that v, u are subwords of an element r of R with $r = vu$ and $l_S(v) > l_S(u)$ (this list is finite). Given a word w on S, we apply the following procedure:

1. If w is empty, stop - the corresponding element is trivial.
2. If w is non empty, compute a cyclically reduced conjugate w' of w - it is at most as long as w, and it represents the trivial element in G iff w does.
3. List all subwords of w', and check whether for one of these subwords v, there exists u such that $(v, u) \in L$.
4. If not, by definition of a Dehn's presentation this means that w', hence w, does not represent the trivial element in G - stop.
5. If there is such a u, and if $w' = w_1vw_2$ for w_1, w_2 subwords of w', then $w' =_G w_1u^{-1}w_2$. Compute the reduced word w'' representing the element $w_1u^{-1}w_2$: it is strictly shorter than w', and represents the trivial element in G iff w' does.
6. Return to step 1. replacing w by w''.

Since the length of the word we apply the loop to strictly decreases with each iteration, the process must stop. $\square$

Theorem 4.31: *A hyperbolic group admits a Dehn's presentation.*

In fact, the converse is also true (we won't prove this).
Given Proposition 4.30, we immediately get that hyperbolic group have solvable word problem.
To prove Theorem 4.31, we will need the two following lemmas

Lemma 4.32: *Let X be a δ-hyperbolic metric space, and let $k \geq 8\delta$. Let $\gamma : [a, b] \to X$ be a k-local geodesic segment, i.e. for any $s, t \in [a, b]$ with $|t - s| < k$ we have $d(\gamma(s), \gamma(t)) = |t - s|$. Then if c is a geodesic segment joining $\gamma(a)$ to $\gamma(b)$, we have*

$$\mathrm{Im}(\gamma) \subseteq B_{2\delta}(\mathrm{Im}(c))$$

Proof. Let $x = \gamma(t)$ realise the maximal distance from $\mathrm{Im}(c)$. Assume first that we can choose y, z points on either side of x such that $d(y, z) < k$ so that γ is a geodesic segment between y and z, but $d(y, x), d(x, z) > 4\delta$. Let y', z' be points of $\imath(c)$ closest to y, z respectively. By cutting the quadrilateral y, z, z', y' diagonally into two geodesic triangles, we see that there exists a point w on one of the sides other than $[y, z]$ such that $d(x, w) < 2\delta$.

If w is on y, y' we get a contradiction: indeed we have $d(x, y') \leq d(x, w) + d(w, y') < 2\delta + d(w, y')$ while

$$\begin{aligned} d(y, y') &= d(y, w) + d(w, y') \geq [d(y, x) - d(x, w)] + d(w, y') \\ &> [4\delta - 2\delta] + d(w, y') \geq 2\delta + d(w, y) \\ &\geq d(x, y') \end{aligned}$$

which contradicts the choice of x. Similarly, w cannot be on $[z, z']$. Thus it is on $[y', z'] \subseteq \mathrm{Im}(c)$ and we are done. If no such y', z' can be chosen, we apply a similar argument (exercise). $\square$

From this we get the following result, which can be thought of as saying that a loop in a hyperbolic space always has a "shortcut" between two points which are at bounded distance apart in the loop.

Lemma 4.33: *Let X be a δ-hyperbolic space. Let $\gamma : [a, b] \to X$ be a non-constant loop in X. Then γ is not a 8δ-local geodesic.*

Proof. If $|b-a| \leq 8\delta$, then taking $s=a, t=b$ is enough since $d(\gamma(a), \gamma(b)) = 0 < |b-a|$. We can thus assume that $|b-a| > 8\delta$.

Suppose that γ is a 8δ-local geodesic segment - by Lemma 4.32, we know that $\mathrm{Im}(\gamma)$ remains within the 2δ neighborhood of the constant geodesic at $\gamma(0)$. Now $d(\gamma(0), \gamma(3\delta)) = 3\delta$ since γ is a 8δ-local geodesic - but this gives a contradiction. □

We can now prove that hyperbolic groups admit a Dehn presentation.

Proof. Let $\pi : F(S) \to G$ be the canonical surjective map. Denote by R_0 the set of words in S such that $r \in \mathrm{Ker}(\pi)$, r is cyclically reduced, and there are subwords u, v of r such that

- r=uv ;
- $l_S(v) \leq 8\delta + 1$;
- $d(1, \pi(u)) = l_S(u)$;
- $l_S(u) < l_S(v)$.

Note that the condition $d(1, \pi(u)) = l_S(u)$ implies that any path labeled by the word u is a geodesic segment. Take R to be the set of all cyclically reduced conjugates of elements of R_0, so that $\langle S \mid R \rangle$ is a symmetric presentation.

Let w be a nontrivial word in S such that $\pi(w) = 1$. We want to show it can be written as a product of conjugates of elements of R (so that the presentation $\langle S \mid R \rangle$ is indeed a presentation for G), and that there exists a subword of w which represents more than half of a relator (to show that it is in fact a Dehn presentation). We prove both by induction on the length n of w.

The word w labels a cycle $\gamma : [0, n] \to X(G, S)$ in the Cayley graph, where $n = l_S(w)$ and we assume that the parametrization is such that γ sends $[j-1, j]$ isometrically onto the j-th edge. By Lemma 4.33, we see that we can find k, l integers in $[0, n]$ with $|l-k| \leq 8\delta + 1$ such that $\gamma |_{[k,l]}$ is not a geodesic segment. By our choice of parametrization, we must have $d(\gamma(k), \gamma(l)) < |l-k|$. Denote by v the label of the path $\gamma |_{[k,l]}$, and by u the label of a geodesic segment between $\gamma(l)$ and $\gamma(k)$. We can assume without loss of generality that the concatenation of $[\gamma(l), \gamma(k)]$ and $\gamma |_{[k,l]}$ has no backtrack (otherwise we could chose l, k closer).

We see that v is a subword of w, that $w = w_1 v w_2$ for subwords w_1, w_2 of w, and that uv labels a cycle in the Cayley graph, so that $\pi(uv) = 1$. By our no backtrack hypothesis, uv is cyclically reduced. Since $l_S(v) \leq 8\delta + 1$, and $l_S(u) < l_S(v)$, $r = uv$ is in R.

If w_1, w_2 are empty this concludes the proof. If not, by induction hypothesis $w_1 u^{-1} w_2$ (which is strictly shorter than w) can be written as a product of conjugates of elements of R, hence so can $w = (w_1 u^{-1} w_2) w_2^{-1} (uv) w_2$. This finishes the proof. □

Note that hyperbolic groups also have solvable conjugacy problem, and Sela showed that the isomorphism problem is solvable for torsion-free hyperbolic groups.

4.5 Subgroups of hyperbolic groups

Is a subgroup of a hyperbolic group always hyperbolic?

This cannot be true in such generality. First, there might be some non finitely generated subgroups, for example, we saw that the free group on infinitely many generators can be seen as a subgroup of $\mathbb{F}_2$. There are also examples of finitely generated but non finitely presented subgroups which embed in hyperbolic groups (via the Rips construction); and even of finitely presented but non hyperbolic subgroups of hyperbolic groups (harder, due to Noel Brady).

Can a hyperbolic group contain a $\mathbb{Z}^2$? A priori, it could contain a twisted copy (i.e. a non quasi-isometrically embedded copy). But in fact, this cannot be. This comes from the following facts

Theorem 4.34: *Let g be an element of a hyperbolic group G. Then the centralizer $C(g)$ of g is quasiisometrically embedded, and if g has infinite order $\langle g \rangle$ has finite index in $C(g)$.*

If G contained a copy of $\mathbb{Z}^2$, the centralizer of any of its elements should contain $\mathbb{Z}^2$, which would contradict its being virtually cyclic.

Note that another consequence of the theorem is that if any two elements of G commute, they generate a virtually cyclic group. If they don't commute, what do they generate?

In free groups:

Remark 4.35: *Let a, b be two elements of a free group $\mathbb{F}$: they generate a subgroup of $\mathbb{F}$, which must be free, and of rank at most 2. If a, b do not commute, the rank is exactly 2.*

In fact, it is possible to show:

Proposition 4.36: *Sufficiently high powers of infinite order noncommuting elements in a hyperbolic group generate a free subgroup.*

In fact, conjecture: a group is hyperbolic iff it has finite K(1) and contains no Baumslag-Solitar subgroup.

5 Limit groups

Another class of "geometrically defined" groups. Hyperbolic groups were close to free groups because their Cayley graphs are almost trees. Here is another class of groups which are close to free groups, in another sense.

5.1 The space of marked groups

Definition 5.1: *A marked group is a pair (G, S) where G is a group and $S = (s_1, \dots, s_k)$ is an ordered generating set for G.*

Two marked groups $(G, (s_1, \dots, s_k))$ and $(G', (s'_1, \dots, s'_{k'}))$ are identified if $k = k'$ and the bijection $s_i \mapsto s'_i$ extends to an isomorphism.

The set of all (isomorphism classes of) marked groups (G, S) where S is a k-tuple is denoted $\mathcal{G}_k$.

Note that if G is a group and T, S are distinct generating sets, then (G, S) and (G, T) are not in general equal as marked groups.

Exercise 5.2: Show that $(\mathbb{Z}, 1)$ and $(\mathbb{Z}, -1)$ are isomorphic as marked groups (and thus identified in $\mathcal{G}_\infty$). Show that $(\mathbb{Z}, (2, 3))$ and $(\mathbb{Z}, (1, 3))$ are not.

Here are two other ways to think about marked groups:

Remark 5.3:
- *a marked group is a group G together with an epimorphism $\pi : \mathbb{F}_k \to G$ (if $a_1, \dots, a_k$ standard basis of $\mathbb{F}_k$, the marking S is given by $s_i = \pi(a_i)$).*
- *choosing a point in $\mathcal{G}_k$ corresponds exactly to choosing a normal subgroup in $\mathbb{F}_k$.*

We want to say that two marked groups are close if their generators satisfy the same relations of a given length:

Definition 5.4: *Let (G, S) and (G', S') be two points in $\mathcal{G}_k$. Let*

$$R((G,S),(G',S')) = \max\{n \mid \forall w \text{ reduced word on } k \text{ letters with } l(w) \leq n, \\ w(S) =_G 1 \iff w(S') =_{G'} 1\}$$

The space of marked groups is the set $\mathcal{G}_k$ endowed with the metric d defined by:

$$d((G,S),(G',S')) = 2^{-R((G,S),(G',S'))}$$

Exercise 5.5: Check this is a metric

So (G, S) and (G', S') are at least 2^{-r}-close, iff they satisfy exactly the same relations of length at most r.

Geometrically:

Exercise 5.6: $R((G, S), (G', S')) \geq r$ iff the balls of radius $r/2$ of their Cayley graphs are isomorphic as labeled graphs (that is, there is a graph isomorphism between them which sends edges labeled s_i to edges labeled s'_i)

Examples of convergent sequences:

Example 5.7:
- the sequence $(\mathbb{Z}/m, (1))$ converges to $(\mathbb{Z}, (1))$ as m tends to ∞.

  Indeed, $R((\mathbb{Z}/m, (1)), (\mathbb{Z}, (1))) \geq m - 1$ since in $(\mathbb{Z}/m, (1))$ there are no relations of length less than m;

- the sequence $(\mathbb{Z}, (1, 2m))$ converges to $\mathbb{Z}^2$ with the standard generating set as m tends to ∞.

  Indeed, $R((\mathbb{Z}, (1, 2m)), (\mathbb{Z}, (1))) \geq 2m$ (in $(\mathbb{Z}, (1, 2m))$ aside from the relations induced by commutation of the form $a^k b^j a^{-k} b^{-j}$, the shortest relation is $a^{2m} b$ which has length $2m + 1$).

By a similar argument it can be shown that $\mathbb{Z}^n$ can be obtained as a limit of some marking of $\mathbb{Z}$.

Proposition 5.8: *The metric space $\mathcal{G}_k$ is compact.*

Proof. $\mathcal{G}_k = \{K \trianglelefteq \mathbb{F}_k\} \subseteq \{0,1\}^{\mathbb{F}_k}$ which is compact by Tychonov (the topology on $\mathcal{G}_k$ is indeed the topology induced by the product topology on $\{0,1\}^{\mathbb{F}_k}$: two marked groups are close if the kernels of the corresponding quotient maps agree on a large finite number of elements).

Thus it is enough to show that $\mathcal{G}_k$ is closed. Let K_n be a sequence of normal subgroups of $\mathbb{F}_k$ and suppose K_n converges to a subset K of $\mathbb{F}_k$ in the product topology. We need to show $K \in \mathcal{G}_k$, that is, that K is a normal subgroup of $\mathbb{F}_k$.

Let $k, k' \in K$. For all n large enough, both k and k' are in K_n. Now K_n is a subgroup of $\mathbb{F}_k$ so $k^{-1}k'$ lies in K_n for all n large enough. Thus it also lies in K.

Similarly, if $k \in K$ and $h \in \mathbb{F}_k$, for all n large enough $k \in K_n$ hence $hkh^{-1} \in K_n$ by normality. Thus $hkh^{-1} \in K$. □

Proposition 5.9: *The set $\mathcal{A} = \{(G,S) \in \mathcal{G}_k \mid G$ is abelian $\}$ is both open and closed.*

Proof. Let $(G,S) \in \mathcal{A}$. Then any group (G',S') at distance less than 2^{-4} is abelian, indeed then for all i,j we have $s'_i s'_j (s'_i)^{-1}(s'_j)^{-1} = 1$.

Suppose that $(G_n, S_n) \to (G,S)$ and G_n abelian for all n. For n large enough $d((G_n,S_n),(G,S)) < 2^{-4}$ so (G,S) satisfies the same relations of length 4 as (G_n,S_n) so (G,S) is abelian. □

In a similar way we can show:

Proposition 5.10: *Let ϕ be a universal formula in the language of groups. The set $\mathcal{U}_\phi = \{(G,S) \in \mathcal{G}_k \mid G \models \phi\}$ is closed.*

Here we are talking about first-order formulas - the language of groups is the set of symbols $\mathcal{L} = (\cdot, ^{-1}, 1)$. A universal formula in this language is a formula (which is equivalent to a formula) of the form

$$\forall x_1 \dots x_p \bigvee_{i=1}^{N} \bigwedge_{j=1}^{M} w_{i,j}(x_1, \dots, x_p) = (\neq) 1.$$

An existential formula is equivalent to a formula of the form

$$\exists x_1 \dots x_p \bigvee_{i=1}^{N} \bigwedge_{j=1}^{M} w_{i,j}(x_1, \dots, x_p) = (\neq) 1.$$

Proof. Suppose that $(G_n,S_n) \to (G,S)$. Suppose $G \not\models \phi$: we can find witnesses $g_1, \dots g_p \in G$ such that none of the conjunctions $\bigwedge_{j=1}^{M} w_{i,j}(g_1, \dots, g_p) = (\neq)1$ hold. The g_i can be seen as words $\tilde{g}_i(S)$ in S.

Let R be larger than the lengths of all the $w_{i,j}(\tilde{g}_1(S), \dots, \tilde{g}_p(S))$ seen as words in S.

For n large enough (G_n,S_n) and (G,S) satisfy exactly the same relations of length R, hence $\tilde{g}_1(S_n), \dots, \tilde{g}_p(S_n)$ in G_n witness the fact that $G_n \not\models \phi$. □

Remark 5.11: *In particular recover that $\mathcal{A}$ is closed since $\mathcal{A} = \mathcal{U}_\phi$ for $\phi : \forall x \forall y \; xy = yx$.*

But cannot extend the openness to the general case: in abelian case, you know that if a group fails to satisfy ϕ, can find "witnesses" of length 1, in general the length of these witnesses is arbitrary.

5.2 Limit groups

Definition 5.12: *We define $\mathcal{L}_k$ to be the closure in $\mathcal{G}_k$ of the set*

$$\mathcal{F} = \{(G,S) \mid G \text{ is free }\}$$

Caution! do not require of G to be free on S.

Definition 5.13: *We say that G is a limit group if there exists an integer k and a marking $S = (s_1, \dots, s_k)$ such that $(G,S) \in \mathcal{L}_k$.*

Exercise 5.14: If G is a limit group, then for any marking S of G, there exists a sequence (G_n, S_n) converging to (G, S) with G_n free.

Exercise 5.15: Show that every finitely generated subgroup of a limit group is a limit group.

Example 5.16:
- Free groups are limit groups;
- Free abelian groups are limit groups (limits of $\mathbb{Z}$).

First properties of limit groups:

Proposition 5.17:
- *Limit groups are torsion free;*
- *Limit groups are commutative transitive;*
- *Any two elements in a limit group which do not commute generate a free group of rank* 2.

Proof. By the proposition above, any universal formula satisfied by free groups is also satisfied by limit groups.

- Fix n. The following formula holds in any free group: $\forall x\ x = 1 \vee x^n \neq 1)$, thus it holds in all limit groups.
- All free groups satisfy $\forall x, y, z\{y \neq 1 \wedge [x, y] = 1 \wedge [y, z] = 1\} \to [x, z] = 1$, hence so does any limit group.
- True in free groups by Remark 4.35. Thus for any non empty reduced word w on two elements the formula $\phi_w : \forall x, y[x, y] \neq 1 \to w(x, y) \neq 1$ holds in $\mathbb{F}$, hence in any limit group. Thus if a, b are elements in a limit group which do not commute, no non trivial word on a, b represents the trivial element, hence a and b generate a free group of rank 2.

□

Example 5.18: The group $\mathbb{F}_2 \times \mathbb{Z}$ is NOT a limit group, since it is not commutative transitive

5.3 Equivalent characterizations of limit groups

5.3.1 Universal theory

Proposition 5.19: *Let G be a fg group. Then G is a non abelian limit group iff it has the same universal theory as $\mathbb{F}_2$.*

Remark 5.20:
1. *All non abelian free groups have the same universal theory. Indeed, for any $k > 1$ we have that $\mathbb{F}_2 \leq \mathbb{F}_k$ so $\mathrm{Th}_\forall(F_k) \subseteq \mathrm{Th}_\forall(\mathbb{F}_2)$, and $\mathbb{F}_k$ embeds in $\mathbb{F}_2$ so the other inclusion also holds.*
2. *Let G, G' be groups. Then $\mathrm{Th}_\forall(G) = \mathrm{Th}_\forall(G')$ iff $\mathrm{Th}_\exists(G) = \mathrm{Th}_\exists(G')$. (This is because the negation of a universal formula is an existential formula).*

Proof. Suppose G is a non abelian limit group: it contains two noncommuting elements, hence it contains a copy of $\mathbb{F}_2$, hence $\mathrm{Th}_\forall(G) \subseteq \mathrm{Th}_\forall(\mathbb{F}_2)$. On the other hand, if ϕ is a universal formula satisfied by all free groups, it will be satisfied by G since this is a closed property.

Suppose G is a fg group which has the same universal theory as $\mathbb{F}_2$. Let $S = (s_1, \dots, s_k)$ be a finite generating set for G. For each N, write the following formula:

$$\phi_N : \exists x_1, \dots, x_k \bigwedge_{w \in B_N(\mathbb{F}_k)} w(x_1, \dots, x_k) = (\neq) 1$$

where we put $=$ if $w(s_1, \dots, s_k) =_G 1$ and $\neq$ otherwise. This holds in G, hence it holds in $\mathbb{F}_2$ (if not, its negation, which is a universal formula, would hold in $\mathbb{F}_2$). Let $S(n) = (s_1(n), \dots, s_k(n))$ be witnesses that this holds. It is easy to see that $(\mathbb{F}_2, S(n))$ converges to (G, S), so G is a limit group. □

5.3.2 Fully residually free

Definition 5.21: *Say that a group G is residually free if for any nontrivial element g, there exists a morphism $f : G \to \mathbb{F}$ where $\mathbb{F}$ is a free group such that $f(g) \neq 1$.*

Say that a group is fully residually free if for any finite set of non trivial elements $g_1, \dots, g_q$, there is a morphism $f : G \to \mathbb{F}$ such that for all i we have $f(g_i) \neq 1$.

Lemma 5.22: *Let G be a fg fully residually free group. Then G is limit .*

Proof. Suppose G is fully residually free, and generated by a finite set S. For each $n \in \mathbb{N}$, let $f_n : G \to \mathbb{F}(n)$ be a homomorphism which does not kill the non trivial elements of S-word length at most n. Let $S_n = f_n(S)$ and let G_n be the (free) subgroup of $\mathbb{F}(n)$ generated by S_n. Clearly for any word w on S, if $w(S) = 1$ in G then $w(S_n) = f_n(w(S)) = 1$ in G_n, and for any w of length at most n, if $w(S) \neq 1$ then $w(S_n) \neq 1$. Thus $(G_n, S_n) \to (G, S)$. □

In fact, the converse is also true so we have:

Proposition 5.23: *Let G be a fg group. Then G is limit iff it is fully residually free.*

But this is harder to prove. Let's do it.

Remark 5.24: *If we know that the limit group G admits a finite presentation $\langle S \mid r_1(S), \dots, r_q(S)\rangle$, then it is easy: indeed, let m be the maximum of the length of the r_i's and of the lengths of the elements that we wish to preserve, and let (G', S') be a marked free group which satisfies exactly the same relations of length m as (G, S). Then the map $G \to G'$ given by $S \mapsto S'$ extends to a homomorphism, since all the relations of the presentation of G hold in G', and this homomorphism does not kill any of the elements we wanted to preserve.*

But we do not know that G is fp. In fact, it is true but very hard to show that all limit groups are in fact fp.

We will use the following fact

Lemma 5.25: *Free groups are equationally Noetherian, namely for any (possibly infinite) system $\Sigma(x_1, \dots, x_k)$ of equations (i.e. expressions $w(x_1, \dots, x_k) = 1$ where w reduced word), there exists a finite subsystem $\Sigma_0(x_1, \dots, x_k)$ such that a tuple $(a_1, \dots, a_k)$ in a free group satisfies Σ iff it satisfies Σ_0.*

Remark 5.26: *This is true of equations over fields: Recall that the ring of polynomials $K[X_1, \dots, X_n]$ is Noetherian, that is, there are no infinite ascending chains of ideals. In particular if Σ is an infinite set of polynomials $P_1(X_1, \dots, X_n), P_2(X_1, \dots, X_n), \dots$, if we define $I_j = (P_1, \dots, P_j)$ we have that for some m, the ideal I_m contains all of Σ. In particular all of the equations $P_j = 0$ for $j > m$ are "consequences" of the first m equations.*

This means precisely that the system of equations Σ is equivalent to the finite subsystem $P_1 = 0, \dots, P_m = 0$.

Proof. We saw that $\mathbb{F}$ embeds in $SL_2(\mathbb{R})$ which we can see as a subvariety of $\mathbb{R}^4$. Each equation $w(x_1, \dots, x_k) = 1$ in $\mathbb{F}$ translates as 4 polynomial equations in the coefficients of the x_i's viewed as elements of $\mathbb{R}^4$.

The set $V_\Sigma = \{(a_1, \dots, a_k) \in \mathbb{R}^{4k} \mid \Sigma(a_1, \dots, a_k)\}$ is a subvariety of $\mathbb{R}^{4k}$, it is the intersection of the decreasing sequence of varieties $V_{\Sigma_p} = \{(a_1, \dots, a_k) \in \mathbb{R}^{4k} \mid \Sigma_p(a_1, \dots, a_k)\}$ where Σ_p consists of the first p equations.

By Noetherianity of the polynomial rings on $\mathbb{R}$, $V_\Sigma = V_{\Sigma_p}$ for some p, so $V_\Sigma \cap \mathbb{F}^k = V_{\Sigma_p} \cap \mathbb{F}^k$ which proves the claim. □

We can now prove the proposition

Proof. Suppose $G = \langle S \mid r_1(S), r_2(S), \dots\rangle$. By equational Noetherianity of the free group, there exists p such that for any $S' \in \mathbb{F}$, $r_i(S') = 1$ for all i iff $r_i(S) = 1$ for $i \leq p$. We can thus proceed with the proof as in the case where G is fp. □

5.4 Morphisms to the free group

We didn't get to this part eventually but I leave it for those of you who might be interested...

We show the following result

Proposition 5.27: *Let G be a finitely presented group. There exists finitely many limit quotients $\eta_i : G \to L_i$ of L such that any morphism $f : G \to L_i$ factors through one of the morphisms η_i.*

Note that if G itself is limit, the result is trivial.

Remark 5.28: *Let $\langle s_1, \ldots, s_k \mid \Sigma(s_1, \ldots, s_k)\rangle$ be a presentation for G. The set $\mathrm{Hom}(G, \mathbb{F})$ of all morphisms from G to a free group can be thought of as the set of solution to the system of equation $\Sigma(x_1, \ldots, x_k)$ in $\mathbb{F}$ - understanding morphisms from G to free groups is thus a way of understanding equations over free groups.*

Note also that we may in fact fix the rank of $\mathbb{F}$ to 2, since any finitely generated free group embeds in $\mathbb{F}_2$ (in fact we even know that the image of a morphism from G to a free group is a free subgroup of rank at most k).

Consider a quotient $p : G \to G'$ between finitely generated groups. Any morphism $f : G' \to \mathbb{F}$ gives a morphism $f \circ p : G \to \mathbb{F}$, and by surjectivity of p if f_1, f_2 are distinct then so are $f_1 \circ p$ and $f_2 \circ p$. Thus $\mathrm{Hom}(G', \mathbb{F}) \hookrightarrow \mathrm{Hom}(G, \mathbb{F})$.

Lemma 5.29: *Consider a sequence of surjective morphisms $G_1 \to G_2 \to G_3 \to \ldots$. The corresponding sequence $\mathrm{Hom}(G_1, \mathbb{F}) \hookleftarrow \mathrm{Hom}(G_2, \mathbb{F}) \hookleftarrow \mathrm{Hom}(G_3, \mathbb{F})\ldots$ must stabilize (i.e. the inclusions are all equalities for i large enough)*

Proof. We choose a generating set $s_1, \ldots, s_k$ for G_1, and take the images of the s_i as generating sets for each G_i (we abuse notation and still denote them by s_i). Let $\langle s_1, \ldots, s_k \mid \Sigma_i(s_1, \ldots, s_k)\rangle$ be a presentation for G_i : we may assume that the system Σ_i is properly contained in the system Σ_{i+1}. The set $\mathrm{Hom}(G_i, \mathbb{F})$ can be thought of as the set of solution to the system of equation $\Sigma_i(x_1, \ldots, x_k)$ in $\mathbb{F}$ - by equational Noetherianity, the system $\bigcup_i \Sigma_i$ is equivalent to a finite subsystem, thus there is q such that for all $i \geq q$, the sets of solutions to Σ_i in $\mathbb{F}$ is the same as the set of solutions to Σ_q - in other words, the sequence $\mathrm{Hom}(G_i, \mathbb{F}) \hookleftarrow \mathrm{Hom}(G_{i+1}, \mathbb{F})$ stabilizes. $\square$

To prove the proposition, we introduce the following notion

Definition 5.30: *Let H be a group. We denote by $\mathrm{Hom}(H, \mathbb{F})$ the set of homomorphisms from H to a free group. The residually free quotient of H is the group*

$$RF(H) = H / \bigcap_{f \in \mathrm{Hom}(H, \mathbb{F})} \mathrm{Ker} f$$

We denote the quotient map by $\pi : H \to RF(H)$.

Remark 5.31:

- *Any homomorphism from H to a free group factors through $RF(H)$: indeed, the elements in $\mathrm{Ker}\pi = \bigcap_{f \in \mathrm{Hom}(H,\mathbb{F})} \mathrm{Ker} f$ are in the kernel of all the morphisms from H to a free group.*
- *The residually free quotient $RF(H)$ is residually free: let $h \in RF(H) - \{1\}$, there is a non trivial element $h' \in H$ such that $\pi(h') = h$. Since h' is not in $\mathrm{Ker}\pi$, it means that there is a morphism $f : H \to \mathbb{F}$ such that $f(h') \neq 1$. Now $f = \bar{f} \circ \pi$ for some morphism $\bar{f} : RF(H) \to \mathbb{F}$ which satisfies $\bar{f}(h) \neq 1$.*
- *Consider a proper quotient $p : R \to R'$ between residually free groups. The inclusion $\mathrm{Hom}(R', \mathbb{F}) \hookrightarrow \mathrm{Hom}(R, \mathbb{F})$ is strict: if $r \in \mathrm{Ker} p - \{1\}$, there is a morphism $f : R \to \mathbb{F}$ such that $f(r) \neq 1$ - this morphism cannot come from a morphism $R' \to \mathbb{F}$.*

We can now prove the proposition

Proof. As noted, we may assume that G is not a limit group, in particular it is not fully residually free. This means that there exists a finite set of non-trivial elements $\{g_1, \ldots, g_r\}$ such that for any morphism $f : G \to \mathbb{F}$, we have $f(g_i) = 1$ for some i. Thus we get finitely many proper quotients $p_i : G \to Q_i = G/\langle\langle g_i \rangle\rangle$ through which any element of $\mathrm{Hom}(G, \mathbb{F})$ factors, but the Q_i are not necessarily limit groups.

Any morphism from Q_i to a free group factors through $RF(Q_i)$ - it is still true that any element of $\mathrm{Hom}(G, \mathbb{F})$ factors through one of the $RF(Q_i)$. Now for each $RF(Q_i)$ which is not a limit group, we may proceed as we did with G to find finitely many proper quotients through which any morphism $RF(Q_i) \to \mathbb{F}$ factors. We keep going in this way until we get to limit group.

This has to stop, otherwise we get an infinite sequence of proper epimorphisms between residually free groups $R_1 \to R_2 \to R_3 \to \ldots$: the corresponding sequence of embeddings $\mathrm{Hom}(G_1, \mathbb{F}) \hookleftarrow \mathrm{Hom}(G_2, \mathbb{F}) \hookleftarrow \mathrm{Hom}(G_3, \mathbb{F}) \ldots$ must stabilize by Lemma 5.29, but by our remark above the inclusions are all strict. $\square$

From this last remark, together with Lemma 5.29, we immediately get

Corollary 5.32: *Any sequence of proper quotients*

$$R_1 \twoheadrightarrow R_2 \twoheadrightarrow R_3 \twoheadrightarrow \ldots$$

between residually free groups R_i is finite.

Proposition 5.27 is in fact a compactness result, as shown by the following alternative proof. We will use the fact (true, but hard) that all limit groups are in fact finitely presented.

Definition 5.33: *If (G, S) is a marked group, say thats (G', S') is a marked quotient of (G, S) if G' is a quotient of G via an epimorphism $\pi : G \to G'$ which sends S to S'. In other words, (G', S') is a marked quotient of (G, S) iff any relation satisfied by the elements of S is satisfied by the elements of S'.*

Denote by $\mathcal{G}(G, S)$ the set of marked quotients of (G, S).

Remark 5.34:
- *$\mathcal{G}(G, S)$ is a closed set: if $(G_n, S_n) \to (G', S')$, any relation satisfied by S in G is satisfied by S_n in G_n for all n, hence it is satisfied by S' in G'.*
- *Let $(G'', S''), (G', S') \in \mathcal{G}(G, S)$. Then (G'', S'') is a marked quotient of (G, S) iff the quotient map $G \to G''$ factors through the quotient map $G \to G'$.*
- *If G admits a finite presentation $\langle S \mid R \rangle$, then (G, S) admits a neighborhood U in $\mathcal{G}_k$ which consists entirely of marked quotients of (G, S), i.e. $U \subseteq \mathcal{G}(G, S)$.*

Proof. Consider the set $\mathcal{L}(G, S) = \mathcal{G}(G, S) \cap \mathcal{L}_k$ of marked limit quotients of (G, S). It is a closed subset of $\mathcal{G}_k$ hence it is compact.

As we said, any marked group in $\mathcal{L}(G, S)$ is finitely presented, hence it admits a neighborhood which consists entirely of quotients of itself. By compactness, there is a finite cover of $\mathcal{L}(G, S)$ by neighborhoods $U(L_1, S_1), \ldots, U(L_m, S_m)$.

Let now $f : G \to \mathbb{F}$ be a morphism to a free group, without loss of generality we may assume f surjective. Hence $(\mathbb{F}, f(S))$ is a marked quotient of (G, S), in fact it is a marked limit quotient of (G, S). In particular, it lies in one of the neighborhoods $U(L_i, S_i)$. This implies that $(\mathbb{F}, f(S))$ is a marked quotient of (L_i, S_i), in other words, that the quotient map $f : G \to \mathbb{F}$ factors through the morphism $p_i : G \to L_i$. $\square$

Proposition 5.27 is a first step in giving a full parametrization of the set of homomorphisms from a given finitely generated group to a free group. Indeed we have

Theorem 5.35: *(Makanin-Razborov diagram) Let G be a finitely generated group. There exists a finite tree T with root vertex G such that*

- *every vertex other than G is a limit group;*

- *every edge from a vertex L to one of its sons L' is an epimorphism $\pi : L \to L'$;*
- *every leaf is a free group;*

and such that for any morphism f from G to some free group $\mathbb{F}$, there exists

- *a branch in the tree* $G \xrightarrow{\pi_1} L^1 \xrightarrow{\pi_2} L^2 \xrightarrow{\pi_2} \dots \xrightarrow{\pi_r} L^r = F$
- *a sequence $\sigma_1, \dots, \sigma)_r$ with $\sigma_i \in \mathrm{Aut}(L^i)$;*
- *an injective morphism $j : F \hookrightarrow \mathbb{F}$;*

such that

$$f = j \circ \sigma^r \circ \pi^r \circ \sigma_{r-1} \circ \dots \circ \sigma_1 \circ \pi_1$$

The tree is called the Makanin-Razborov diagram for G. As noted, Proposition 5.27 gives the first floor of this diagram. The fact that branches are finite is a consequence of Corollary 5.32.