Lecture notes on Geometric Group Theory

Under construction

Alessandro Sisto

July 3, 2014

Contents

Chapter 1

Foreword

These are lecture notes for the 2014 course on Geometric Group Theory at ETH Zurich.

Geometric Group Theory is the art of studying groups without using algebra. Here's a rather effective description from Ric Wade:

[Geometric Group Theory] is about using geometry (i.e. drawing pictures) to help us understand groups, which can otherwise be fairly dry algebraic objects (i.e. a bunch of letters on a piece of paper).

The way to use geometry to study groups is considering their (isometric) actions on metric spaces. Many theorems in Geometric Group Theory look like: *Let G be a group acting "nicely" on a "nice" space. Then G ...*

The core part of the course is devoted to (Gromov-)hyperbolic spaces and groups.

I'm experimenting a bit with the style. What I want to do is explaining concepts, ideas, etc. in a way that resembles how you would explain them in person more than a traditional book/ set of lecture notes. "Traditional" books and lecture notes about Geometric Group Theory or hyperbolic groups include the following:

- *A course on geometric group theory*, by Bowditch
- *Metric spaces of non-positive curvature*, by Bridson and Haefliger
- *Les groupes hyperboliques de Gromov*, by Coornaert, Delzant and Papadopoulos
- *Sur les groupes hyperboliques d'aprés Mikhael Gromov*, by Ghys and de la Harpe

The references will all be given at the end.

Part I

Cayley graphs, quasi-isometries and Milnor-Švarc

Chapter 2

The Cayley graph

The aim of this chapter is to introduce a metric space called Cayley graph that can be naturally associated to a finitely generated group (together with a fixed finite generating set). The group acts on its the Cayley graph in a natural way.

Notation: In this chapter G will always denote a group generated by the *finite* set $S \subseteq G \backslash \{1\}$. For convenience we also assume $S = S^{-1} = \{s^{-1} | s \in S\}$.

There's no deep reason to require $1 \notin S$, but there are a few points where allowing 1 as a generator makes things more annoying to write down. Requiring $S = S^{-1}$ is also not very important but sometimes convenient.

2.1 Metric graphs

This section can be safely skipped if you know what a metric graph is. Or even if you can just guess it.

Recall that a graph Γ consists of points called *vertices* and copies of $[0, 1]$ connecting pairs of vertices called *edges*. Also, interiors of distinct edges are disjoint. (Sometimes one requires that there are no double edges, but we don't need to.)

Suppose that we assigned to each edge e of a given connected graph Γ some positive number $l(e)$ (its length). Then we can define on Γ a pseudo-metric, which we now describe in two ways.

If we regard each edge as an isometric copy of $[0, l(e)]$, we have a natural way of defining the length of a path consisting of the concatenation of finitely many subpaths of edges. We can then define the "distance" $d(x, y)$ between two points $x, y \in \Gamma$ to just be the infimum of the lengths of paths as above connecting them. This infimum might be 0, and this is the only reason why d may fail to define a metric. The infimum is never 0 for $x \neq y$ if there is a lower bound on the length of the edges.

Here is another way to describe d. For each edge e fix a homeomorphism $\phi_e : e \to [0, 1]$ as in the definition of edge. Define the auxiliary function ρ

in the following way. If x, y belong to the same edge e, then define $\rho(x,y) = l(e)|\phi_e(x) - \phi_e(y)|$, and set $\rho(x,y) = +\infty$ otherwise. Finally, set

$$d(x,y) = \inf_{x=x_0,\ldots,x_n=y} \sum \rho(x_i, x_{i+1}).$$

$\{x_i\}$ as above is usually called *chain* (from x to y).

Lemma 2.1.1. *In the definition one can equivalently only take chains $x = x_0, \ldots, x_n = y$ with the additional constraint that x_i is a vertex for $i \neq 0, n$.*

Proof. If $\sum \rho(x_i, x_{i+1})$ is finite then for any x_i which is not a vertex both x_{i-1} and x_{i+1} have to be contained in the same edge as x_i, if $i \neq 0, n$. Removing x_i from the chain does not increase the value of the sum because $\rho(x_{i-1}, x_{i+1}) \leq \rho(x_{i-1}, x_i) + \rho(x_i, x_{i+1})$ by triangular inequality. Hence, starting from any chain, we can iteratively remove the non-vertices in the "middle" part and find a new chain $\{y_i\}$ satisfying our extra requirement and so that $\sum \rho(y_i, y_{i+1}) \leq \sum \rho(x_i, x_{i+1})$. Hence, the infimum taken over the smaller set of chains that we are considering coincides with the the infimum over all chains from x to y. $\square$

We will mostly use edges of length 1, but occasionally edges of different lengths will also show up.

2.2 Definition and examples

Here is the definition of Cayley graph.

Definition 2.2.1. The *Cayley graph $Cay(G,S)$* of G with respect to S is the metric graph with

1. vertex set G,
2. an edge connecting $g, h \in G$ if and only if $g^{-1}h \in S$, i.e. if and only if there exists $s \in S$ with $h = gs$,
3. all edges of length 1.

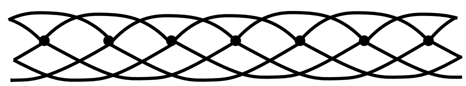
We denote the metric on $Cay(G,S)$ as d_S.

2.2.1 Examples

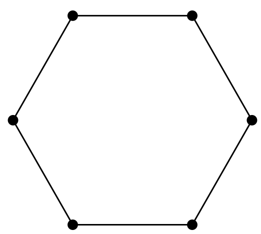
Here are some Cayley graphs that are easy to describe and draw.

1. $Cay(\mathbb{Z}, \{\pm 1\})$ is isometric to $\mathbb{R}$:

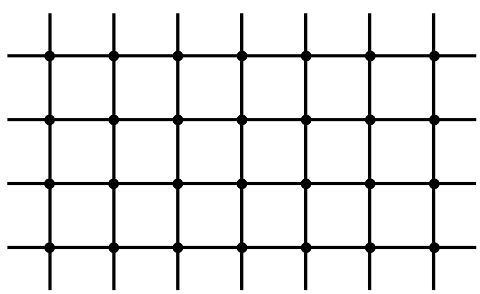
2. Changing the generating set does change the Cayley graphs. For example, $Cay(\mathbb{Z}, \{\pm 2, \pm 3\})$ looks like this:



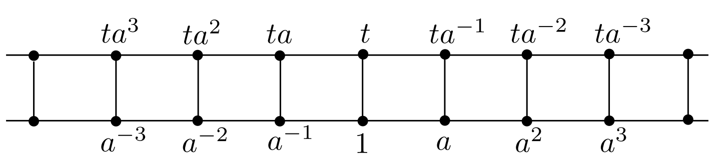
3. $Cay(\mathbb{Z}/n, \{\pm 1\})$ (with $n \geq 3$) is an n-gon, and it is isometric to a rescaled copy of S^1 with the arc-length metric:



4. $Cay(\mathbb{Z}^2, \{\pm(0,1), \pm(1,0)\})$ is the "grid" in $\mathbb{R}^2$:

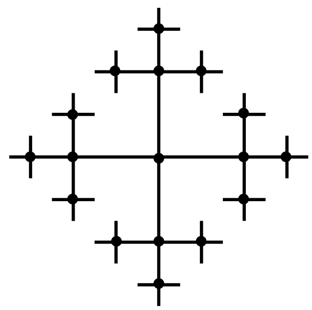


5. $Cay(D_\infty, \{a^{\pm 1}, t\})$, where $D_\infty = \langle a, t | t^2 = 1, tat = a^{-1} \rangle$ is the infinite dihedral group, is a bi-infinite ladder:



Notice that the powers of a increase from left to right in the bottom part but decrease from left to right in the top part.

6. $Cay(F_2, \{a^{\pm 1}, b^{\pm 1}\})$ (where a, b are a basis of the free group on two generators F_2) is what's called a tree and looks like this:



In the rest of this chapter we will explore the properties of Cayley graphs. Here is the first one.

Fact 1: For $g, h \in G$, we have $d_S(g,h) = \min\{n | \exists s_1, \ldots, s_n \quad g^{-1}h = s_1 \ldots s_n\}$ (and $d_S(g,h) = 0$ if $g = h$).

In words, the distance between g and h is the minimum length of a word in the alphabet S representing $g^{-1}h$, i.e. the *word length* of $g^{-1}h$. One often denotes $d_S(1,g)$ by $|g|_S$.

Fact 1 follows directly from Lemma 2.1.1.

2.3 G acts on the Cayley graph

Fact 2: G acts by isometries on $Cay(G,S)$. Such action extends the action of G on itself by left translation (i.e. $g(h) = gh$).

In order to convince ourselves of Fact 2 notice that, for any $g, h_1, h_2 \in G$, there is an edge from h_1 to h_2 if and only if there is an edge between gh_1 and gh_2. This is just because $(gh_1)^{-1}gh_2 = h_1^{-1}h_2$. If you prefer (I do), if you obtain h_2 from h_1 by multiplying on the right by some $s \in S$, i.e. $h_2 = h_1 s$, then clearly you also obtain gh_2 from gh_1 multiplying on the right by the same s.

Now, using the observation above we can extend the left multiplication by $g \in G$ across the edges of the Cayley graph.

More formally, in order to define an action by isometries of G on $Cay(G,S)$ one has to assign to each $g \in G$ an isometry ϕ_g. Such isometry can be written down as follows. For $x \in G$, $\phi_g(x)$ is just gx. For x on the edge from, say, h_1 to h_2, $\phi_g(x)$ is the only point on the edge from gh_1 to gh_2 satisfying $d_S(gh_1, \phi_g(x)) = d_S(h_1, x)$.

The following properties have to be checked for $g \mapsto \phi_g$ to define an action by isometries, and they are both straightforward.

1. ϕ_g is an isometry,

2. $\phi_{gh} = \phi_g \circ \phi_h$.

From now on, for notational convenience we will write g instead of ϕ_g, that is to say, if you like, we identify the group element g with its induced isometry on the Cayley graph.

Don't read this, it's not worth it:

What we just defined is a left action, meaning that $g \to \phi_g$ is a homomorphism from G to the isometry group of $Cay(G, S)$ IF composition in the said isometry group is defined right-to-left, i.e. $(i_1 \circ i_2)(x) = i_1(i_2(x))$ for all isometries i_1, i_2 and $x \in Cay(G, S)$.

2.4 $Cay(G, S)$ is nice

In this section we address the question: How good is $Cay(G, S)$ as a metric space?

Here is the first good property of Cayley graphs.

Fact 3: $Cay(G, S)$ is a proper metric space, i.e. its closed balls are compact.

Fact 3 is just a consequence of the fact that any ball, say of integer radius, is the union of finitely many edges, and each edge is compact.

Exercise: How many edges can there be at most in a ball of radius n (in terms of the cardinality of S)? What is a pair (group, generating set) where such number of edges is maximal?

You may have noticed that up to now edges have just been an annoyance, and we would have been better off just putting the metric as in Fact 1 on G. But fear not, we are about to use them.

Fact 4: $Cay(G, S)$ is a geodesic metric space.

You can skip the next subsection if you already know what this means.

2.4.1 Geodesic metric spaces

Let us fix a metric space X from now until the end of the subsection.

For α a path in X (i.e. a continuous map $\alpha : [0, 1] \to X$), let us define the *length* of α as

$$l(\alpha) = \sup_{0=t_0 \leq \dots \leq t_n = 1} \sum d(\alpha(t_i), \alpha(t_{i+1})).$$

The idea is rather simple. Suppose you want to formally define the notion of length of a path. One property you would like is that if you approximate your curve by a concatenation of "straight lines" then the length of such concatenation approximates the length of the path (well, if the path has finite length). The length of a "straight line" should be the same as the distance between the endpoints, whence the definition above.

Two further properties you would like are content of the following two remarks.

Remark 2.4.1. For any path α, we have $l(\alpha) \geq d(\alpha(0), \alpha(1))$. In fact, the sums appearing in the definition of length are all greater or equal than $d(\alpha(0), \alpha(1))$ by triangular inequality (and induction).

Remark 2.4.2. Let us denote the concatenation of the paths α, β by $\alpha*\beta$. Then $l(\alpha * \beta) = l(\alpha) + l(\beta)$. To prove the inequality $\leq$, just notice that given a chain of points "approximating" α and one "approximating" β, we can concatenate them and form a chain of points for $\alpha * \beta$. To prove $\geq$, notice that given a chain of points "approximating" $\alpha * \beta$ we can add a point and make it a concatenation of chains for α and β. If this sounds mysterious, it's probably a good idea to work out the details yourself.

We mentioned "straight lines" above. They are actually called geodesics:

Definition 2.4.3. The path α is a *geodesic* if $l(\alpha) = d(\alpha(0), \alpha(1))$. The metric space X is *geodesic* if for any pair of points of X there is a geodesic connecting them.

Hence, geodesics are the most efficient paths to get between two points. We conclude the subsection with two useful properties of geodesics.

Proposition 2.4.4.

1. *A subpath of a geodesic is a geodesic.*
2. *if α is a geodesic then for any $s \leq t \leq u$ we have*

$$d(\alpha(s), \alpha(t)) + d(\alpha(t), \alpha(u)) = d(\alpha(s), \alpha(u)).$$

The idea behind the first item is just that if a path α is as efficient as possible, i.e. it is a geodesic, then all its subpaths have to be as efficient as possible, for otherwise we could detour a subpath of α and create a shorter path connecting the endpoints of α.

Item 2 says that the triangular inequality is actually an equality "along α" and is a formal way of saying that α behaves like a "straight line". The idea is that if the triangular inequality was not an equality for $\alpha(s), \alpha(t), \alpha(u)$, then it would be more efficient to avoid going through $\alpha(t)$ when getting from $\alpha(s)$ to $\alpha(u)$.

Proof. 1) Suppose that the geodesic α can be written as a concatenation $\beta*\gamma*\delta$. Here is the computation we need, all (in)equalities are explained below. It may look scary, but I promise it's rather straightforward.

$$\begin{aligned} d(\alpha(0), \alpha(1)) = & l(\alpha) = l(\beta) + l(\gamma) + l(\delta) \geq \\ & d(\beta(0), \beta(1)) + d(\gamma(0), \gamma(1)) + d(\delta(0), \delta(1)) \geq \\ & d(\alpha(0), \alpha(1)). \end{aligned}$$

The first equality holds because α is a geodesic and the second one holds by Remark 2.4.2. The first $\geq$ follows from Remark 2.4.1, while the second one from the triangular inequality (using $\alpha(0) = \beta(0)$, $\beta(1) = \gamma(0)$, etc.).

All inequalities have to be equalities because the first and last term are equal, and in particular the only way that the first $\geq$ can be an equality is if $l(\beta) = d(\beta(0), \beta(1))$ and similarly for γ, δ, i.e. if β, γ, δ are geodesics.

2) The subpath β of α from $\alpha(s)$ to $\alpha(u)$ is a geodesic by item 1). Hence

$$d(\alpha(s), \alpha(u)) = l(\beta) \geq d(\alpha(s), \alpha(t)) + d(\alpha(t), \alpha(u)) \geq \\ d(\alpha(s), \alpha(u))$$

The first $\geq$ follows from the fact that $l(\beta)$ is the supremum of certain sums, one of which is the one to the right of $\geq$. The second $\geq$ is a triangular inequality.

Once again, all inequalities have to be equalities, in particular the last one, which is the one we need. □

2.4.2 Back to Cayley graphs

One way of showing the (hopefully very believable) fact that $Cay(G, S)$ is geodesic is the following. First of all, it is not difficult to see from Fact 1 that any two elements of G are joined by a geodesic in $Cay(G, S)$. Another easy case is when we pick two points lying on a common edge. Now, we know that the distance between $x, y \in Cay(G, S)$ is

$$d_S(x, y) = \inf_{x=x_0,\dots,x_n=y} \sum \rho(x_i, x_{i+1}),$$

where $x_i \in G$ for $i \neq 0, 1$ (see Lemma 2.1.1). It is then not difficult to see that the following formula holds for all x, y that do not lie on a common edge:

$$d_S(x, y) = \inf\{d_S(x, g) + d_S(g, h) + d_S(h, y) | d(x, g) < 1, d(h, y) < 1\}.$$

In words, in order to go from x to y one has first to go to an endpoint of an edge containing x, then go to some other vertex and then finally go to y staying on an edge.

The infimum is actually a minimum because there are only at most two g's and two h's satisfying the requirement. If the minimum is realized when considering g, h, then it is readily checked that a concatenation of a geodesic from x to g, one from g to h and one from h to y gives a geodesic from x to y. (Because the lengths of such geodesics are $d_S(x, g), d_S(g, h), d_S(h, y)$ respectively, so the length of the concatenation equals $d_S(x, g) + d_S(g, h) + d_S(h, y) = d_S(x, y)$.)

We will use geodesics all the time, Fact 4 is going to be very convenient. The arguments that we will make with geodesics in $Cay(G, S)$ can presumably all be rephrased in terms of chains of points in G, but they would be way more painful to write down. We're making a little extra effort now to make life easier later.

Here is an overkill to prove that $Cay(G, S)$ is geodesic. The distance on $Cay(G, S)$ is defined as an infimum of lengths of paths. Using Arzela-Ascoli and the fact that $Cay(G, S)$ is proper, one can show that a sequence of paths from x to y whose lengths converge to the distance between x, y converges, provided that the said paths are parametrized by arc length.

2.5 The action of G is nice

So far we can say that we proved that every finitely generated group G acts by isometries on a proper geodesic metric space. Sounds good, doesn't it?

However, it only *sounds* good. Here is another construction of such an action. Take G. Take a metric space X consisting of only one point. Make G act on X trivially (of course). And we don't even need that G is finitely generated!

The message here is that if you have an action you don't just want the space being acted on to be nice, you also want the action itself to be nice. Hence, we now address the question: How good is the action $G \curvearrowright Cay(G, S)$?

Fact 5: The action of G on $Cay(G, S)$ is proper.

We say that an action of the group G on the metric space X is proper if for any $x \in X$ and any ball $B \subseteq X$ there are only finitely many elements of G that map x inside B. It is easy to check that this holds, keeping into account that the orbit of a vertex of the Cayley graph is naturally identified with G itself. The idea is that orbit points have to be "well-spaced" and leave every compact set as you move away from the identity in G.

Here are two straightforward consequences of properness that are good to keep in mind. If the action of G on X is proper then

1. stabilizers of points are finite, and

2. orbits do not have accumulation points.

Fact 6: The action of G on $Cay(G, S)$ is cobounded.

The action of G on the metric space X is said to be cobounded if there is a ball $B \subseteq X$ whose G-translates cover the whole X, that is to say $G \cdot B = X$. Another way of saying this is: There is a point $x \in X$ and some constant R so that any point in X is within distance R from a point in the orbit of X.

While properness is about having not too many orbit points in a confined space, coboundedness is about orbits points being pretty much everywhere.

It is easy to see that one can take a ball of radius, say, 1 in the case of Cayley graphs.

Let us now sum up what we did so far.

Theorem 2.5.1. *Every finitely generated group acts properly and coboundedly by isometries on a proper, geodesic metric space. An example of such an action is the natural action on a Cayley graph.*

2.6 A relaxing exercise

Many concepts have been introduced so far, so it is a good time to see them in action.

Exercise 1. If the group G acts properly and coboundedly by isometries on $\mathbb{R}$, then it contains a finite index subgroup isomorphic to $\mathbb{Z}$.

Exercise 2. $\mathbb{Z}^2$ can act faithfully on $\mathbb{R}$ with dense orbits.

The hint for Exercise 2 is to make $(a, b) \in \mathbb{Z}^2$ act as the translation by $a + b\sqrt{2}$.

Here is a detailed outline of Exercise 1. You are very welcome to try and solve it rather than reading the solution, of course...

Suppose that the action of G is given by the homomorphism $\Psi : G \to Isom(\mathbb{R})$, the group of isometries of $\mathbb{R}$. First of all, G contains a subgroup G' of index at most 2 so that each element acts on $\mathbb{R}$ as a translation. In fact, any isometry of $\mathbb{R}$ either preserves or reverses the order, and in the first case the isometry is a translation. The map from $Isom(\mathbb{R})$ to $\mathbb{Z}/2$ that maps the isometry ϕ to the non-trivial element of $\mathbb{Z}/2$ if and only if ϕ reverses the order is a homomorphism. If K is the kernel of such map, $G' = \Psi^{-1}(K \cap \Psi(G))$ has the required properties.

Ok, now let us set $m = \inf(G' \cdot 0 \cap \mathbb{R}_{>0})$, the infimum of the "positive part" of the orbit of 0.

First of all, from the fact that the action is cobounded, one sees that $G' \cdot 0 \cap \mathbb{R}_{>0}$ is non-empty. Secondly, the infimum must actually be a minimum, because orbits cannot have accumulation points. For the same reason, m is strictly positive.

Let us now consider the map $\varphi : G' \to \mathbb{R}$ so that $\varphi(g) = \Psi(g)(0)$. It is not hard to see that the image of φ is actually contained in $m\mathbb{Z}$. More importantly, we claim that φ is a homomorphism.

In fact, let us denote by t_g, for $g \in G'$, the real number so that $\Psi(g)(x) = x + t_g$ for each $x \in \mathbb{R}$ (remember, G' acts by translations). In particular, $\varphi(g) = t_g$. We can deduce that $t_{gh} = t_g + t_h$ from the fact that Ψ defines an action: $t_{gh} = \Psi(gh)(0) = \Psi(g)(\Psi(h))(0) = t_g + t_h$.

Finally, the kernel F of φ is the stabiliser of 0, which is finite by properness. Hence, we have the exact sequence

$$1 \to F \to G' \to m\mathbb{Z} \approx \mathbb{Z} \to 1.$$

Whenever we have such a sequence there is always a section $s : m\mathbb{Z} \to G'$, i.e. a homomorphism $s : m\mathbb{Z} \to G'$ so that $(\varphi \circ s) = id$. In particular, $s(m\mathbb{Z})$ is isomorphic to $\mathbb{Z}$, and has finite index in G', whence in G. We finally found the finite index subgroup of G isomorphic to $\mathbb{Z}$, as required.

Chapter 3

Quasi-isometries

Very often, you want to study a group rather than a pair group/generating set. However, constructing a Cayley graph requires fixing a finite generating set, and we don't like this.

In this chapter we answer the question: To what extent does the Cayley graph of a given group depend on the generating set?

The answer requires the notion of quasi-isometry.

Definition 3.0.1. Let X, Y be metric spaces and let $f : X \to Y$ be a map from X to Y. We say that f is a (K, C)-*quasi-isometric embedding* if for any $x, y \in X$ we have

$$\frac{d(x,y)}{K} - C \leq d(f(x), f(y)) \leq K d(x, y) + C.$$

The (K, C)-quasi-isometric embedding f is a (K, C)-*quasi-isometry* if for any $y \in Y$ there is some $x \in X$ with $d(f(x), y) \leq C$ (i.e. f is *coarsely surjective*).

A quasi-isometric embedding is just a (K, C)-quasi-isometric embedding for some K, C, and similarly for quasi-isometries.

Notice that a $(K, 0)$-quasi-isometric embedding is just a bi-Lipschitz map. Hence, a good way of thinking about quasi-isometric embeddings is that they are bi-Lipschitz maps at a large scale. Another useful heuristic to keep in mind is that quasi-isometric embeddings "don't distort distances too much".

Notice that when you have a (K, C)-quasi-isometric embedding you get no information at all at scales below C, and in particular a quasi-isometric embedding need not be continuous. It is as if you had a(n infinite) ruler with marks spaced by C. If C is, say 1 km, it is pointless to try and measure bacteria with it, but if you want to measure galaxies then it's more than adequate. In this spirit, coarse surjectivity is the right replacement for surjectivity in our setting because we cannot measure whether $f(x)$ actually coincides with y or it's just C-close to it.

Remark 3.0.2. The first inequality in the definition of quasi-isometric embedding can be rewritten as $d(x, y) \leq K d(f(x), f(y)) + KC$.

3.0.1 Examples

1. For $v, b \in \mathbb{R}^2$, the map $t \mapsto tv + b$ from $\mathbb{R}$ to $\mathbb{R}^2$ is a quasi-isometric embedding.

2. The map $t \mapsto t^2$ from $\mathbb{R}$ to $\mathbb{R}$ is not a quasi-isometric embedding. The second inequality is the one that fails.

3. The map $t \mapsto \sqrt{t}$ from $\mathbb{R}$ to $\mathbb{R}$ is not a quasi-isometric embedding. This time the first inequality fails.

4. The logarithmic spiral $\mathbb{R}^+ \to \mathbb{R}^2 \approx \mathbb{C}$, which is given by $t \mapsto te^{i\pi \ln t}$, is a quasi-isometric embedding. (Just take derivatives to see that it's Lipschitz. How can you see the lower bound?) However, it's not a quasi-isometry.

5. $Cay(\mathbb{Z}^2, \{\pm(0,1), \pm(1,0)\})$ can be embedded in a natural way into $\mathbb{R}^2$. Such embedding is a quasi-isometry.

3.0.2 Quasi-inverses

Let $f : X \to Y$ be a map between metric spaces. We say that $g : Y \to X$ is a *quasi-inverse* of f if there exists D so that for each $x \in X$ we have $d_X((g \circ f)(x), x) \leq D$ and, similarly, for each $y \in Y$ we have $d_Y((f \circ g)(y), y) \leq D$.

So, a quasi-inverse is just an inverse "up to bounded error".

We record a few useful properties of quasi-isometric embeddings and quasi-isometries.

Proposition 3.0.3. *1. Composition of quasi-isometric embeddings (resp. quasi-isometries) is a quasi-isometric embedding (resp. quasi-isometry).*

2. Let f be a quasi-isometric embedding. Then f is a quasi-isometry $\iff$ it has a quasi-inverse. Also, the quasi-inverse is a quasi-isometry as well.

3. Being quasi-isometric is an equivalence relation.

The first item just says that a composition of maps that don't distort distances too much doesn't distort distances too much as well, and composition of maps that are "surjective up to bounded error" is "surjective up to bounded error".

The idea for the second item is simple as well. We just want to define $g(y)$ to be some $x \in X$ that gets mapped close to y (the coarse version of what one does to define the inverse of a surjective map). This has to be a quasi-inverse, and it cannot distort distances too much because f doesn't.

Proof. 1) Let $f : X \to Y$ and $g : Y \to Z$ be (K, C)-quasi-isometric embeddings. Then for each $x, y \in X$ we have

$$\begin{aligned} d(g(f(x)), g(f(y))) \leq & K d(f(x), f(y)) + C \leq \\ & K^2 d(x, y) + KC + C. \end{aligned}$$

This is one of the inequalities we need to show that $g \circ f$ is a quasi-isometric embedding. The other one can be proven similarly.

Now suppose that f, g are (K, C)-quasi-isometries, and pick $z \in Z$. We want to find $x \in X$ so that $(g \circ f)(x)$ is close to z. We know that there exists $y \in Y$ with $d(g(y), z) \leq C$ and $x \in X$ with $d(f(x), y) \leq C$.

Hence, we get

$$d((g \circ f)(x), z) \leq d(g(f(x)), g(y)) + d(g(y), z) \leq \\ (KC + C) + C,$$

as required.

2) The implication $\Leftarrow$ is straightforward. We need to show that for each y there exists x so that $f(x)$ is close to y. Such x is just $g(y)$, where g is a quasi-inverse of f.

Let us prove the implication $\Rightarrow$. Suppose that f is a (K, C)-quasi-isometry. Define $g(y)$ to be some $x \in X$ so that $d(f(x), y) \leq C$. A good picture to keep in mind is that x is in the preimage of $B_C(y)$, which is some blob of bounded diameter.

By definition, $d(f(g(y)), y) \leq C$ for each y. Let us bound $d(g(f(x)), x)$ for $x \in X$. If we map x across using f and then back using g, we end up in a blob containing x, and in particular within bounded distance from x. Here is the formal estimate:

$$d(g(f(x)), x) \leq Kd(f(g(f(x))), f(x)) + KC \leq 2KC.$$

The first inequality follows from Remark 3.0.2, while the second one from the fact that $f \circ g$ is C-close to the identity.

Let us show that a quasi-inverse g of f is also a quasi-isometry. By what we proved so far, it is enough to show that it is a quasi-isometric embedding. The idea is just that g cannot distort distances too much, otherwise f would have to as well. Let D be as in the definition of quasi-inverse. Here is one inequality, the other one is similar.

$$d(g(y_1), g(y_2)) \leq Kd(f(g(y_1)), f(g(y_2))) + KC \leq Kd(y_1, y_2) + 2KD + KC.$$

The second inequality just follows from $f(g(y_i))$ being D-close to y_i, so that $d(f(g(y_1)), f(g(y_2)))$ is within $2D$ from $d(y_1, y_2)$. If you don't believe it, denote $f(g(y_i))$ by z_i and check these out:

$$d(z_1, z_2) \leq d(z_1, y_1) + d(y_1, y_2) + d(y_2, z_2) \leq d(y_1, y_2) + 2D,$$

$$d(y_1, y_2) \leq d(y_1, z_1) + d(z_1, z_2) + d(z_2, y_2) \leq d(z_1, z_2) + 2D.$$

3) There clearly is a quasi-isometry from any metric space to itself. Transitivity is statement 1), and symmetry follows from 2). $\square$

3.0.3 Cayley graphs and quasi-isometries

And now we are ready to show that "the" Cayley graph of a given group is well-defined up to quasi-isometry.

Proposition 3.0.4. *Let G be a group and S, S' two finite symmetric generating sets for G. Then the identity $id : G \to G$ extends to a quasi-isometry $Cay(G,S) \to Cay(G,S')$.*

Proof. Let us first reduce to considering the vertex sets of the Cayley graphs. Consider the composition

$$Cay(G,S) \xrightarrow{\psi} (G, d_S) \xrightarrow{id} (G, d_{S'}) \xrightarrow{\iota} Cay(G,S'),$$

where ψ is any map mapping $x \in Cay(G,S)$ to some $g \in G$ with $d_S(x,g) \leq 1/2$ and ι is just the inclusion.

Notice that ψ and ι are $(1,1)$-quasi-isometries, so the overall composition is a quasi-isometry if $id : (G, d_S) \to (G, d_{S'})$ is.

The identity is surjective, and we are about to check that it is bi-Lipschitz, which will conclude the proof.

Recall that $d_S(1,g)$ is denoted $|g|_S$ and is the minimal number of generators from S needed to write g (and similarly for S').

Set

$$M = \max\{|x'|_S, |x|_{S'} : x \in S, x' \in S'\}.$$

Now, if $d_S(g,h) = k$, then we can write $g^{-1}h = s_1 \dots s_k$, with $s_i \in S$. What we can do now is "expand" each s_i using the s_i''s to write $g^{-1}h$ as a product of generators from S'. Unfortunately this looks a bit ugly:

$$s_1 \dots s_k = (s'_{1,1} \dots s'_{1,M_1}) \dots (s'_{k,1} \dots s'_{k,M_k}),$$

for some $M_i \leq M$ and $s'_{i,j} \in S'$. So, we have

$$d_{S'}(g,h) = |g^{-1}h|_{S'} \leq Mk \leq M d_S(g,h).$$

The inequality $d_S \leq M d_{S'}$ follows using the same argument. □

3.0.4 Quasi-isometric groups and (un)distorted subgroups

We can now talk about "the" Cayley graph of a group, if we keep in mind that it's well-defined only up to quasi-isometry, and we can also talk about quasi-isometric groups (meaning that they have quasi-isometric Cayley graphs) and groups quasi-isometric to metric spaces.

The following facts are easy to see:

1. An isomorphism of groups is a quasi-isometry (formally: induces a quasi-isometry of Cayley graphs).

2. If H is a subgroup of G, then the inclusion is a quasi-isometry if and only if H has finite index in G.

3. A surjective homomorphism $G \to H$ is a quasi-isometry if and only if the kernel is finite.

Passing to a finite index subgroup or modding out a finite normal subgroup should be seen as "finite perturbations" that cannot be seen from the point of view of quasi-isometries.

Inspired by 2), one may wonder whether the inclusion of a (finitely generated) subgroup in the ambient group is always a quasi-isometric embedding. Unfortunately, this is not the case. Such is life.

Subgroups whose inclusions are a quasi-isometric embeddings are called *undistorted*, the other ones are called *distorted*. Here are some examples, that are going to give us a good excuse to introduce a couple of interesting groups.

1. Any subgroup of an abelian group is undistorted. Exercise.

2. We will see that any cyclic subgroup of a hyperbolic group is undistorted. This is not true for all subgroups, however.

3. The subgroup generated by a in $BS(1,2) = \langle a, t | tat^{-1} = a^2 \rangle$ is isomorphic to $\mathbb{Z}$ and distorted (while the one generated by t is undistorted).

4. The subgroup generated by z in the Heisenberg group $\langle x, y, z | [x, y] = z, [x, z] = [y, z] = 1 \rangle$ is isomorphic to $\mathbb{Z}$ and distorted.

Let us elaborate a bit on item 3). Let us take for granted that a has infinite order. Now, it is easy to inductively see that

$$a^{2^n} = t^n a t^{-n}.$$

For example, $t^2 a t^{-2} = t(tat^{-1})t^{-1} = ta^2t^{-1} = tat^{-1}tat^{-1} = a^4$.

In particular, with respect to the generating set given above, $d(1, a^{2^n}) \leq 2n + 1$. The exponent of a and the distance are then definitely not linearly related, and hence $\langle a \rangle$ is (exponentially) distorted.

By the way, BS stands for Baumslag-Solitar, and $BS(1,2)$ is a Baumslag-Solitar group. The groups $BS(m, n)$ are defined as $\langle a, t | ta^m t^{-1} = a^n \rangle$, and they are a good source of (counter)examples.

Let us now analyse item 3), once again taking for granted that the order of z is infinite. Let us start from $x^n y^n$. Suppose that we want to move all y's to the left of the x's. Let us start from the leftmost y. From $[x, y] = z$ we see that we can switch it with the rightmost x creating a z. Also, z commutes with everything, so we can just move it to the right. Now, we can repeat and move our y one step further to the left, once again creating a z. Repeating this, we get

$$x^n y^n = y x^n y^{n-1} z^n = \cdots = y^n x^n z^{n^2}.$$

So, $z^{n^2} = x^{-n}y^{-n}x^n y^n$, and hence $d(1, z^{n^2}) \leq 4n$. This shows that $\langle z \rangle$ is (quadratically) distorted.

Now, a few words on the Heisenberg group. Another way of describing it is as the group of the 3-by-3 upper-triangular matrices with 1's on the diagonal:

$$\begin{pmatrix} 1 & \mathbb{Z} & \mathbb{Z} \\ 0 & 1 & \mathbb{Z} \\ 0 & 0 & 1 \end{pmatrix}$$

The generators written above are the following ones:

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

Notice that the Heisenberg group is nilpotent, and actually one of the simplest non-abelian nilpotent groups. Also, it is the fundamental group of a 3-manifold, constructed in the following way. Take a torus $T = S^1 \times S^1$, and consider $T \times [0,1]$. Now, given a homeomorphism $\phi : (T \times \{0\}) \to (T \times \{1\})$, one can construct the so-called mapping torus of ϕ by identifying $T \times \{0\}$ and $T \times \{1\}$ via ϕ. For a suitable choice of ϕ, the resulting manifold has fundamental group isomorphic to the Heisenberg group. Can you describe one such ϕ?

3.1 The final exercise

Exercise. Suppose the the finitely generated group G is quasi-isometric to $\mathbb{R}$. Then G is virtually $\mathbb{Z}$.

Chapter 4

Milnor-Švarc Lemma

In this chapter we put together several of the concepts that we have seen so far. The following Theorem tells us that when you have a group acting nicely on a geodesic metric space, then the Cayley graph of your group looks like the space being acted on. It is sometimes called the fundamental lemma of Geometric Group Theory, and it is probably the main reason why one might wish to study groups up to quasi-isometry.

Theorem 4.0.1. *(Milnor-Švarc Lemma) Suppose that the group G acts properly and coboundedly on the geodesic metric space X. Then*

1. *G is finitely generated*
2. *$Cay(G)$ is quasi-isometric to X, via the map[1] $g \mapsto gx_0$ for any given choice of $x_0 \in X$.*

One of the motivating examples of proper cobounded actions arises in the following way. Let M be a compact, connected Riemannian manifold. Its universal cover $\widetilde{M}$ is in a natural way also a Riemannian manifold, and $\pi_1(M)$ acts on it by isometries. Such action is proper and cobounded, so $\pi_1(M)$ is quasi-isometric to $\widetilde{M}$. In particular, one can use Milnor-Švarc Lemma to put restrictions on the kinds of Riemannian metrics that a given manifold can carry. More on this later...

Proof. Fix $x_0 \in X$, and let R be so that every $x \in X$ is within distance R from gx_0 for some $g \in G$. Such R exists because the action is cobounded. Now, consider the subset of G of all elements that map x_0 within distance $2R+1$ from itself. In formulas:

$$S = \{g \in G : d(x_0, gx_0) \leq 2R + 1\}.$$

[1] The map is defined only on the vertex set, so formally one should extend it. No big deal, right?

Notice that S is finite because the action is proper. (S contains the identity and so it does not satisfy our standing assumptions on generating sets. We are going to ignore this.)

Ok, we are now ready for the core of the proof.

- S generates G. Also, $|g|_S \leq d_X(x_0, gx_0) + 2$ for each $g \in G$.

The idea is the following. We can connect x_0 to gx_0 by a geodesic. Such geodesic has a halo of orbit points around it, and we can select a chain made of such points. Two consecutive points will be not too far from each other, and hence the corresponding group elements are going to differ by multiplication by some $s \in S$. This means that our element g is a product of elements of s. Also, we can choose a chain containing a number of points comparable to the length of the geodesic, and this shows the required inequality.

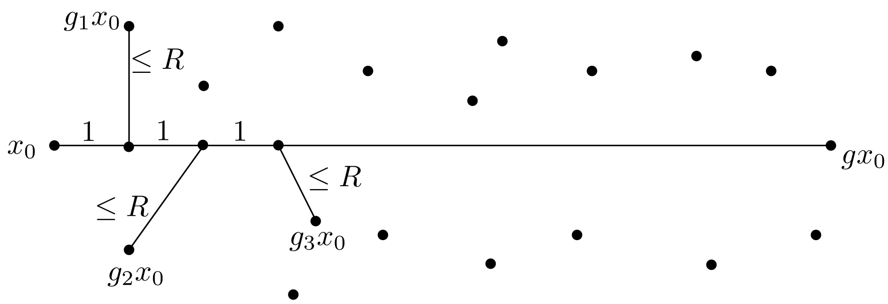


Figure 4.1: From a geodesic in X to a chain of elements of G.

Let's turn this handwavy argument into a proof. Pick a geodesic γ from x_0 to gx_0. We have a sequence of points $x_0 = p_0, \ldots, p_n = gx_0$ so that $d(p_i, p_{i+1}) \leq 1$ and $n \leq d(x_0, gx_0) + 2$ (say all consecutive points are 1 apart except possibly for the last two that are ≤ 1 apart). Now, each p_i is R-close to some $g_i x_0$ (with $g_n = g$ and $g_0 = 1$). Next, we show that $g_{i+1} = g_i s$ for some $s \in S$. Notice that:

$$d(x_0, g_i^{-1} g_{i+1} x_0) = d(g_i x_0, g_{i+1} x_0) \leq 2R + 1.$$

This means, by definition, that $g_i^{-1} g_{i+1}$ is in S, i.e. $g_{i+1} = g_i s_i$ for some $s_i \in S$ as we wanted.

Almost there. We have

$$s_0 \ldots s_{n-1} = (1 s_0) \ldots s_n = (g_1 s_1) \ldots s_{n-1} = \cdots = g_{n-1} s_{n-1} = g_n = g.$$

So, we wrote an arbitrary $g \in G$ as a product of elements of S. Also, we used $n \leq d_X(x_0, gx_0) + 2$ of those, which shows $|g|_S \leq d_X(x_0, gx_0) + 2$ as required.

Ready for the next part.

- $d_X(x_0, gx_0) \leq (2R + 1)|g|_S$.

Here is (an interpretation of) what we have done so far. We have our orbit points of G in X, and we decided to (abstractly) connect those that are not too far way from each other, and we got something connected. Now, if we have a way of going from one point of G to another using the connections we created, then this "projects" to a way of connecting the corresponding orbit points in X. This is how we are going to get the required estimate.

Write $g = s_1 \dots s_k$, with $k = |g|_S$. Denote $g_i = s_1 \dots s_i$ (with $g_0 = 1$). Notice that $g_i x_0$ is not far from $g_{i+1} x_0$:

$$d_X(g_i x_0, g_{i+1} x_0) = d_X(x_0, g_i^{-1} g_{i+1} x_0) = d(x_0, s_{i+1} x_0) \leq 2R + 1.$$

So,

$$d_X(x_0, g x_0) \leq \sum d_X(g_i x_0, g_{i+1} x_0) \leq (2R+1)k = (2R+1)|g|_S,$$

as required.

The bullets above (and the fact that the action on X is by isometries) easily imply the inequalities needed to show that $g \mapsto g x_0$ is a quasi-isometric embedding (on the vertex set of $Cay(G, S)$). The image is R-dense by coboundedness, hence we really described a quasi-isometry. □

4.1 A digression on growth

The material in this section is mostly independent from what is going to happen next. The aim is to introduce an interesting and simple-to-define quasi-isometry invariant that we can couple with Milnor-Švarc Lemma to get some corollaries.

Let G be generated by the finite set S. We define $\beta_{G,S}(n)$ simply as the cardinality of the ball in (G, d_S) of radius n, and we call $\beta_{G,S}$ the *growth function* of G with respect to S.

Here are some examples. In each case with respect to the standard generating sets, we have:

- $\beta_{\mathbb{Z},S}(n) = 2n + 1$,
- $\beta_{\mathbb{Z}^2,S}(n) = 2n^2 + 2n + 1$,
- $\beta_{F_m,S}(n) = 2m(2m-1)^{n-1}$, where F_m is the free group on m generators,
- $\beta_{H,S}(n) = \Theta(n^4)$, where H is the Heisenberg group.

The last example is somehow the most surprising one. It is natural to think of the Heisenberg group as a 3-dimensional object, so one might expect that the growth function goes like n^3 rather than n^4. The metaphysical reason for why it really should be of order n^4 is that, while the x and y directions contribute 1 to the growth exponent, the z direction contributes 2 because it is quadratically distorted and hence there are order of n^2 elements in the z direction within distance n of the identity. In order to prove this, you may wish to write elements in the Heisenberg group in an appropriate normal form...

As usual, we are interested in groups, not pairs group/generating set. In order to get rid of the dependence on the generating set, we define (a partial order and) an equivalence relation on the collection of functions.

Let $f, g : \mathbb{N} \to \mathbb{N}$. Write $f \preceq g$ is there exists some constant C so that for each $n \in \mathbb{N}$ we have

$$f(n) \le Cg(Cn + C).$$

One way to think about it is that for n's "at the same scale", $f(n)$ should be "at most of the same order of magnitude" as $g(n)$. Further, we write $f \asymp g$ if $f \preceq g$ and $g \preceq f$, while we write $f \prec g$ if $f \preceq g$ but $g \not\preceq f$.

Here are basic examples, that after all tell you that $\preceq$ is not too bad, except maybe that it doesn't distinguish exponential functions.

- If $0 < a < b$ then $n^a \prec n^b$.
- If $1 < \alpha < \beta$ then $\alpha^n \asymp \beta^n$.
- for each $a > 0$, $n^a \prec 2^n$.

So, at least we can distinguish exponents of polynomial growth, and we can distinguish polynomial functions from exponential functions.

The definition we gave is the right one to make sure that the $\asymp$-class of the growth function does not depend on the generating set. But more is true.

Proposition 4.1.1. *The $\asymp$-class of the growth function is a quasi-isometry invariant of groups.*

Before proving the proposition (which is not that difficult), let us discuss some applications. The original motivation for Milnor to show the Milnor-Švarc Lemma was that he wanted to study growth functions of fundamental groups of Riemannian manifolds. For example, combining Milnor-Švarc Lemma, a slightly improved version of the Proposition, and well-known volume estimates in Riemannian geometry, one (Milnor) can show that the fundamental group of a compact Riemannian manifold with negative sectional curvature has exponential growth. In particular, for example, the 3-manifold whose fundamental group is the Heisenberg group that we mentioned in Subsection 3.0.4 cannot carry such a Riemannian metric, because the growth of the Heisenberg group is polynomial.

Let us now prove the Proposition.

Proof. Let G, H be groups equipped with the finite generating sets S, T, respectively. Let $f : G \to H$ be a (K, C)-quasi-isometric embedding. We need two facts that follow easily from the definition of quasi-isometric embedding.

The first one is that the image of a ball $B^G(1, n)$ of radius n in G is contained in a ball of radius $Kn + C$ in H.

The second one is that the preimage of any element of H contains at most, say, M elements. This is because it is entirely contained in a ball of radius, say, $KC + 1$ (two points further away than that cannot be mapped to the same point).

Combining the two facts we get:

$$\#B^G(1,n) \leq M \cdot \#B^H(1, Kn + C),$$

one of the inequalities we wanted. The other one can be proven in the same way using a quasi-inverse of f. $\square$

When talking about growth, it is impossible not state the following result of Gromov:

Theorem 4.1.2. *A finitely generated group has at most polynomial growth if and only if it is virtually nilpotent.*

In particular, being virtually nilpotent, a purely algebraic property, is a quasi-isometry invariant.

There's also another result that comes to mind when talking about growth. We saw examples of groups with polynomial growth, and it is easy to see that any group containing a free subgroup has exponential growth, so we get examples of exponential growth as well. Is there anything in between? The answer is yes. In fact, Grigorchuck constructed groups that have growth type e^{n^α} for some $0 < \alpha < 1$.

Part II

Hyperbolic spaces and groups

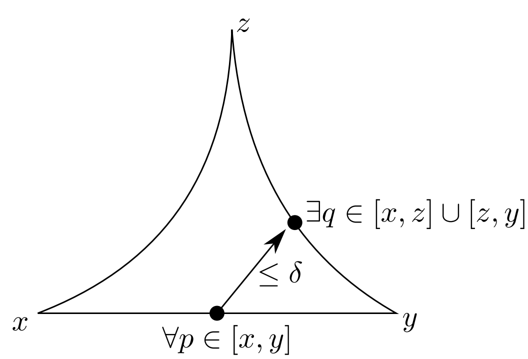
Chapter 5

Definition, examples and quasi-isometry invariance

Hyperbolic spaces are defined by a very simple property of geodesic triangles, but their theory is amazingly rich. Having a sufficiently nice action on a hyperbolic metric space has a lot of consequences.

Definition 5.0.3. The geodesic metric space X is (Gromov-)*hyperbolic* if there exists $\delta \geq 0$ so that for any geodesic triangle $[x,y] \cup [x,z] \cup [z,y]$ and any $p \in [x,y]$ there exists some $q \in [x,z] \cup [z,y]$ with $d(p,q) \leq \delta$. Such δ is called hyperbolicity constant of X.

A triangle satisfying the condition above is called δ-*thin*[1] and looks like this:



The definition can also be reformulated in terms of the notion of neighborhood. For a subset $A \subseteq X$ of the metric space X and $R \geq 0$, the *neighborhood* of A of radius R, denoted $N_R(A)$, is defined as

$$N_R(A) = \{x \in X : d(x, A) \leq R\}.$$

[1] Someone would call it δ-slim.

We can then say that X is hyperbolic if there exists $\delta \geq 0$ so that for any geodesic triangle $[x,y] \cup [x,z] \cup [z,y]$ we have

$$[x,y] \subseteq N_\delta([x,z] \cup [z,y]).$$

We will show that being hyperbolic is a quasi-isometry invariant of geodesic metric spaces. Accepting this for the moment, we can define what it means for a group to be hyperbolic.

Definition 5.0.4. The finitely generated group G is *hyperbolic* if one of the following equivalent conditions hold.

1. G has one hyperbolic Cayley graph.
2. Every Cayley graph of G is hyperbolic.
3. G acts properly and coboundedly on a hyperbolic metric space.

The third condition is equivalent to the other two by Milnor-Švarc Lemma and quasi-isometry invariance of hyperbolicity.

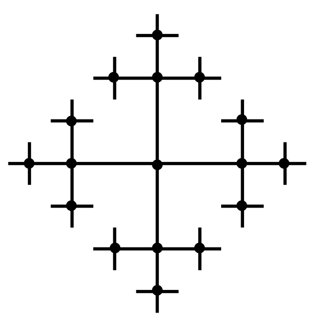
5.1 Examples (for the moment!)

Let us start from the uninteresting examples.

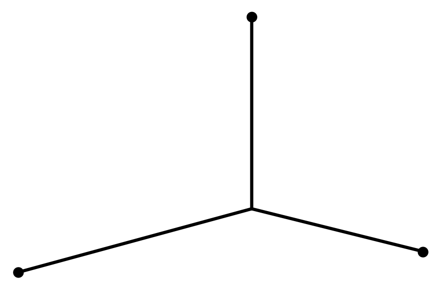
- Metric spaces of bounded diameter are hyperbolic, just take the diameter as δ.
- $\mathbb{R}$ is hyperbolic, as every triangle is degenerate.

And now let us move on to more interesting examples.

- Free groups are hyperbolic. Remember that a free group (on at least two generators) has a Cayley graph that looks like this:



  For such Cayley graph, we can actually take $\delta = 0$, i.e. any side of a geodesic triangle is contained in the union of the other two sides, like in this picture:



- The hyperbolic plane $\mathbb{H}^2$ is hyperbolic (this is where the name comes from), and the same is true for the higher dimensional hyperbolic spaces $\mathbb{H}^n$. Recall that for $g \geq 2$, the connected compact boundary-less surface S_g of genus g admits a Riemannian metric whose universal cover is $\mathbb{H}^2$. In particular, $\pi_1(S_g)$ is hyperbolic.

- More in general, for a compact Riemannian manifold M, we have that $\pi_1(M)$ is hyperbolic if and only if $\widetilde{M}$ is. (Almost) concrete examples of such manifolds in dimension 3 can be constructed as follows. For $g \geq 2$, consider $S_g \times [0,1]$, which has two boundary components homeomorphic to S_g. If we give a homeomorphism of S_g, we can use it to glue the two boundary component and obtain a 3-manifold without boundary. As it turns out, due to a theorem of Thurston, if we choose the gluing map to be a "pseudo-Anosov" (whatever that means), then the manifold M we get has a Riemannian metric with universal cover $\mathbb{H}^3$, so that its fundamental group is hyperbolic. In algebraic terms, the fundamental group can be described by a short exact sequence of the type

$$1 \to \pi_1(S_g) \to \pi_1(M) \to \mathbb{Z} \to 1.$$

We will see more (algebraic/combinatorial) examples later...

5.2 Non-examples

- $\mathbb{R}^2$ and $\mathbb{Z}^2$ are not hyperbolic. This is easy.

- More in general, any group containing a copy of $\mathbb{Z}^2$ is not hyperbolic. We'll see this.

- $BS(1,2)$ and the Heisenberg group are not hyperbolic, for example because they contain distorted cyclic subgroups, while we will see that this cannot be the case for hyperbolic groups. (The Heisenberg group also contains $\mathbb{Z}^2$.)

5.3 List of properties

The point of this section is to impress you with the amount of stuff that is known about hyperbolic spaces. We will see proofs of several of these facts.

- Hyperbolic groups are finitely presented, i.e. every hyperbolic group has a presentation with finitely many (generators and) relators.
- When given a finite presentation, you can ask yourself how you can determine whether a given word in the generating set represents the trivial element or not. This, a bit surprisingly, is not doable in general. In some other cases, there exist only very slow algorithms to do that. However, for any given presentation of a hyperbolic group there is a *linear time* algorithm to determine whether a given word represents the identity or not.
- Any hyperbolic group has finitely many conjugacy classes of finite subgroups.
- The centraliser of any infinite order element of a hyperbolic group is virtually $\mathbb{Z}$, i.e. almost as small as possible.
- Cyclic subgroups of hyperbolic groups are undistorted.

  There are also some "largeness" properties that only require a very mild additional condition. If the group G is hyperbolic and not virtually cyclic (in particular, not finite), then:
- G contains a copy of the free group F_2 on two generators. In particular, G has exponential growth (i.e., it grows as fast as possible).
- G is SQ-universal, i.e. any countable group embeds in some quotient of G. In particular, G has uncountably many non-isomorphic quotients, because there are uncountably many non-isomorphic finitely generated groups and any given finitely generated group contains only countably many finitely generated subgroups.

5.4 Quasi-isometry invariance

In this section we show the quasi-isometry invariance of hyperbolicity. The notion of hyperbolicity is stated in terms of geodesics, but the image of a geodesic via a quasi-isometry need not be a geodesic. However, we will show that in a suitable sense it is within bounded distance from a geodesic, if the ambient space is hyperbolic.

Definition 5.4.1. A (K,C)-*quasi-geodesic* in the metric space X is a (K,C)-quasi-isometric embedding of some interval in $\mathbb{R}$ into X.

The reason why we gave this definition is because the image of a geodesic γ via a (K, C)-quasi-isometric embedding is a (K, C)-quasi-geodesic[2]

It is convenient to introduce now the notion of Hausdorff distance, which is a sensible way of measuring how different two subsets of a metric space are.

Let A, B be subsets of a given metric space, then their *Hausdorff distance* is defined as

$$d_{Haus}(A, B) = \inf\{R : A \subseteq N_R(B), B \subseteq N_R(A)\}.$$

A good way to think about it is: If every point in A is within distance R from B and viceversa, then we have $d_{Haus}(A, B) \leq R$. And conversely, if $d_{Haus}(A, B) \leq R$ then every point in A is within distance R from B and viceversa.

The informal version of the following statement is that quasi-geodesics in hyperbolic spaces stay close to geodesics.

Proposition 5.4.2. *Let X be hyperbolic. Then for every K, C there exists D so that for any quasi-geodesic ρ and any geodesic $\gamma = [x, y]$ with the same endpoints x, y as ρ, we have that the Hausdorff distance between (the images of) γ and ρ is at most D.*

The analogous statement for $\mathbb{R}^2$ is false. For examples, following the x-axis for a while and then the y-axis gives quasi-geodesics (with uniform constants) that cannot possibly be within uniformly bounded Hausdorff distance from geodesics. Another counterexample is given by the logarithmic spiral that was mentioned among the examples of quasi-isometric embeddings.

And here is the corollary we were aiming for.

Corollary 5.4.3. *Let X, Y be geodesic metric spaces. If there exists a quasi-isometric embedding $f : X \to Y$ and Y is hyperbolic, then so is X.*

In particular, if the geodesic metric spaces X, Y are quasi-isometric, then X is hyperbolic if and only if Y is hyperbolic.

In order to prove the corollary we just need to push forward geodesic triangles in X to quasi-geodesic triangles in Y, use that quasi-geodesic triangles in Y are thin and deduce that triangles in X are thin because f doesn't distort distances too much.

Proof. Let's give name to the constants. Let δ be a hyperbolicity constant for Y and suppose that $f : X \to Y$ is a (K, C)-quasi-isometric embedding and let D be as in Proposition 5.4.2.

Ok, now consider a geodesic triangle $[x, y] \cup [x, z] \cup [z, y]$ in X and $p \in [x, y]$. Then $f(p)$ is D-close to some p_1 on a geodesic from $f(x)$ to $f(y)$. By hyperbolicity and up to switching x and y, p_1 is δ-close to some p_2 on a geodesic

[2]This is slightly imprecise. What is true is that the image of a geodesic *parametrised by arc length* is a quasi-geodesic, where a geodesic parametrised by arc length is a map $\gamma : (I \subseteq \mathbb{R}) \to X$ so that $d(\gamma(t_1), \gamma(t_2)) = |t_1 - t_2|$ for every $t_i \in I$. Any geodesic can be reparametrised to have this property.

from $f(x)$ to $f(z)$. Finally, p_2 is D-close to $f(q)$ for some $q \in [x, z]$. So, we only need to show that p and q are close. This is easy: We have $d(f(p), f(q)) \leq 2D + \delta$, and so, using that f is a (K, C)-quasi-isometric embedding we get $d(p, q) \leq (2D + \delta)K + KC$. □

5.4.1 Proof of Proposition 5.4.2

We will use the following preliminary lemma that basically says that we can replace any quasi-geodesic with a "tamer" one that stays in in a controlled neighborhood of the first one.

Lemma 5.4.4. *For any K, C there exist K', C', D so that the following holds. For any (K, C)-quasi-geodesic γ to a geodesic metric space X there exists a (K', C')-quasi-geodesic γ' so that:*

1. *γ' has the same endpoints as γ,*
2. *$d_{Haus}(\gamma, \gamma') \leq D$.*
3. *for any subpath β of γ, say from a to b, we have $l(\beta) \leq K'd(a, b) + C'$.*

The proof is remarkably tedious and not very interesting, so we will skip it. A detailed argument is presented in [Bridson-Haefliger]. The idea is easy: if the domain of γ is I, you look at $\gamma(I \cap \mathbb{Z})$ and interpolate with geodesics.

Notation. Unless otherwise stated, from now on we fix the notation of Proposition 5.4.2, and δ will denote a hyperbolicity constant for X. Furthermore, we assume that the quasi-geodesic ρ satisfies the conditions of the lemma (with K, C replacing K', C'). In view of the lemma, it is enough to prove the proposition for such quasi-geodesics. We will also assume $d(x, y) \geq 1$.

The proof can be split in three parts.

The logarithmic estimate

For the moment, we show that the distance from any point on the geodesic to the quasi-geodesic is bounded logarithmically in the length of the quasi-geodesic. The content of this part of the proof is the following lemma.

Lemma 5.4.5. *Let X be a hyperbolic space with hyperbolicity constant δ. Suppose that $p \in X$ lies on the geodesic $[x, y]$ and that α is a path from x to y, of length at least 1. Then*

$$d(p, \alpha) \leq \delta \log_2(l(\alpha)) + 2.$$

Proof. We can assume that the length of α is finite.

If $l(\alpha) \leq 2$, then the statement is clear because $d(x, y) \leq 2$ and hence $d(p, \alpha) \leq d(p, x) \leq 2$.

The core of the proof consists in splitting α into two parts of equal length and proceeding inductively, the base case being $l(\alpha) \leq 2$.

Suppose $l(\alpha) \geq 2$. Let $q \in \alpha$ be so that q splits α into two parts α_1, α_2 of equal length. We know that p is within distance δ from some point p' either on a geodesic $[x, q]$ or a geodesic $[q, y]$, let us say that the first case holds and that α_1 is the part of α with endpoints x, q. So, using induction we have the straightforward computation:

$$d(p, \alpha) \leq d(p, p') + d(p', \alpha_1) \leq$$
$$\delta + \delta \log_2 \Big(l(\alpha)/2 \Big) + 2 = \delta \log_2(l(\alpha)) + 2.$$

That's it. □

As the length of the quasi-geodesic ρ is at most $Kd(x, y) + C$ we get the following.

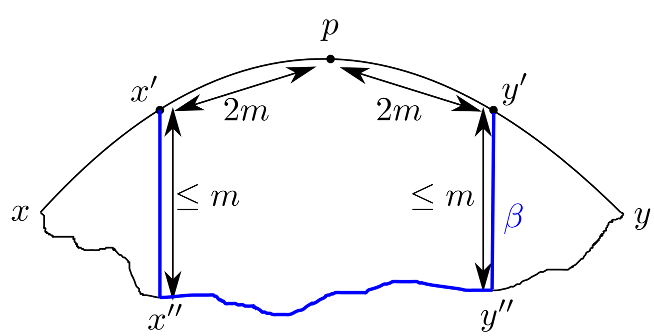
Corollary 5.4.6. *In the notation of Proposition 5.4.2, for any* $p \in [x, y]$ *we have* $d(p, \rho) \leq \delta \log_2(Kd(x, y) + C) + 2$.

Picking the worst point

In this subsection we show that for each $p \in [x, y]$ we have $d(p, \rho) \leq m$, for some m that depends on K, C, δ only.

The strategy we adopt now to improve the logarithmic estimate is picking the "worst point" along $[x, y]$, meaning the one furthest away from the quasi-geodesic ρ, and try to find some configuration that allows us to jump efficiently from p to ρ.

Pick $p \in [x, y]$ with $d(p, \rho) = \max\{d(q, \rho) : q \in [x, y]\} = m$. We now consider points as in the following picture:



Formally, we pick x' along $[x, y]$ and before p with $d(x', x) = \min\{2m, d(p, x)\}$, i.e., either the point before p along $[x, y]$ at distance $2m$ from p and if such point does not exist we take $x' = x$. Also, we define y similarly on the other side. Next, we choose $x'' \in \rho$ with $d(x', x'') \leq m$ (which is possible by definition of m). We take $x'' = x'$ if $x' = x$, and define y'' similarly.

The reason why we moved a distance $2m$ from p is that any point on a geodesic $[x', x'']$ or $[y', y'']$ has distance at least $2m - m = m$ from p. So, if β is the concatenation of $[x', x'']$, a subpath of ρ from x'' to y'' and $[y'', y']$, we have

• $d(p, \beta) \geq m$.

Also, $d(x'', y'') \leq m + 2m + 2m + m = 6m$. So, the length of the bottom part of β (meaning the subpath of ρ) is at most $6mK + C$, and

• $l(\beta) \leq (6K + 2)m + C$.

Putting together the logarithmic estimates and the two properties of β we get:

$$
\begin{aligned}
m \leq & d(p, \beta) \leq \\
& \delta \log_2(l(\beta)) + 2 \leq \\
& \delta \log_2 \Big((6K + 2)m + C \Big) + 2
\end{aligned}
$$

The last term is logarithmic in m, so there is a bound on m that depends on δ, K, C only.

No large detours

We already showed half of what we wanted to prove, namely that any point on $[x, y]$ is close to ρ. We now conclude the proof by showing that any point on ρ is D-close to $[x, y]$, for a suitable D that depends on K, C, δ.

The idea is to bound the length of the subpaths of ρ that make a "detour" outside the neighborhood $N_m([x, y])$, where m is as in the second part of the proof (and it depends on δ, K, C only).

Pick any $q \in \rho$. If $d(q, [x, y]) \leq m$, we are happy. Otherwise, consider the two subpaths ρ_1, ρ_2 of ρ that concatenate at q. By the property of m, any point in $[x, y]$ is m-close to either ρ_1 or ρ_2. One endpoint x of $[x, y]$ is close to ρ_1, while the other one y is close to ρ_2. Travelling along $[x, y]$ we then see that at some point p we switch from being close to ρ_1 to being close to ρ_2, and hence $d(p, q_i) \leq m$ for some points $q_1 \in \rho_1$ and $q_2 \in \rho_2$. The length of the subpath of ρ from q_1 to q_2 (which contains q!) is bounded by $2mK + C$, because its endpoints are within distance $2m$ of each other. In particular

$$d(q, [x, y]) \leq d(q, q_1) + d(q_1, p) \leq 2mK + C + m.$$

As q was any point on ρ, we are done.

Chapter 6

$\mathbb{H}^n$

The aim of the chapter is to describe THE motivating examples of (Gromov-) hyperbolic spaces. Annoyingly, they are called (real) hyperbolic spaces, and they are denoted by $\mathbb{H}^n$, where $n \geq 2$ is an integer.

Some Riemannian geometry is about to show up, but no worries, it will go away soon.

6.1 Two models

One of the possible ways of defining $\mathbb{H}^n$ is "the unique (up to isometry) complete simply connected Riemannian manifold with all sectional curvatures -1 at every point".

Luckily, there are concrete models for $\mathbb{H}^n$, and actually there are several of them. This turns out to be very convenient because different properties of $\mathbb{H}^n$ are clear in different models. We now present two of the models.

The half-space model. Consider an open half-space in $\mathbb{R}^n$, i.e. $\mathbb{R}^{n-1} \times \mathbb{R}_{>0}$, where we denote the first $n-1$ coordinates by x_i and the last one by y. The Riemannian metric on the half space that makes it (isometric to) $\mathbb{H}^n$ is

$$\frac{1}{y^2}\Big(\sum dx_i^2 + dy^2\Big).$$

It is important to note that the metric is a pointwise rescaling of the Euclidean metric (i.e. the two metrics are conformally equivalent) and that it is invariant under translations in the x_i coordinates.

The ball model. Consider now the open unit ball in $\mathbb{R}^n$, and denote the coordinates by x_i. We denote the square of the Euclidean norm of a point $\overline{x} = (x_i)$ by $|\overline{x}|^2 = \sum x_i^2$. The Riemannian metric at the point $\overline{x}$ is in this case

$$\frac{4}{(1-|\overline{x}|^2)^2}\sum dx_i^2.$$

Once again, the metric is a pointwise rescaling of the Euclidean metric, and in this case it is invariant under rotations around the origin.

It can be seen directly that the two models are isometric. The isometry can be described in terms of what goes under the name of inversion across a circle. We won't describe it here, but one good thing about it is the remarkable property that it maps (Euclidean) lines and circles in one model to (Euclidean) lines and circles in the other one.

For concreteness, we henceforth focus on $\mathbb{H}^2$, the hyperbolic plane, but several of the things we are about to say hold in higher dimension as well. In this case, the first model is called half-plane model and the second model is called the disk model or Poincarè disk model. We use coordinates x, y so that the Riemannian metrics of the two models become:

$$\frac{1}{y^2}\Big(dx^2 + dy^2\Big),$$

and

$$\frac{4}{(1-|\overline{x}|^2)^2}\Big(dx^2 + dy^2\Big),$$

respectively.

6.2 Lots of isometries!

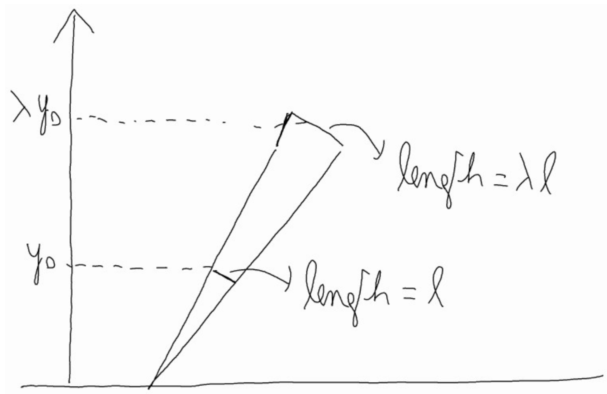
One of the coolest as well as most important properties of $\mathbb{H}^2$ is that it is very symmetric, meaning that it has lots of isometries. Let us describe the isometries that are easy to see in the two models.

6.2.1 In the half-plane model

The following transformations of the half-plane are isometries of $\mathbb{H}^2$. The first two are clear from the formula for the Riemannian metric.

- Translations in the x coordinate, i.e. maps of the type $(x, y) \mapsto (x + t, y)$, for some fixed t.
- Reflections across the y-axis, i.e. $(x, y) \mapsto (-x, y)$.
- Dilations, i.e. maps of type $(x, y) \mapsto (\lambda x, \lambda y)$ for some $\lambda > 0$.

In order to see that dilations are isometries one can just make a computation, and what one gets is that the derivative of the dilation contributes a factor λ^2 to the Riemannian metric, while the $1/y^2$ factor contributes a factor $1/\lambda^2$. More pictorially, if one has a tiny tiny path in the half-plane, its length in the hyperbolic metric is going to be pretty much the Euclidean length times $1/y_0$, where y_0 is the y-coordinate of the path. When hitting the half-plane with the dilation, we multiply the Euclidean length by λ, but we end up somewhere with y-coordinate λy_0, so we need to rescale the new length by λy_0 instead of y_0. So, the two effects cancel out, which means that the dilation preserves lengths of paths computed in the hyperbolic metric and hence the hyperbolic metric itself.



Notice that we already have enough isometries to show that $\mathbb{H}^2$ is homogeneous, i.e. that for any $p, q \in \mathbb{H}^2$ there is an isometry $f : \mathbb{H}^2 \to \mathbb{H}^2$ with $f(p) = q$. (How can one show this?)

6.2.2 In the disk model

The following transformations of the disk are (clearly) isometries of $\mathbb{H}^2$.

- Reflections across diameters.
- Rotations around the origin.

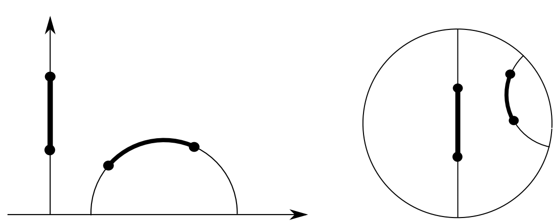
Using these additional isometries one can show that $\mathbb{H}^2$ is actually bi-homogeneous, meaning that for any p_1, p_2, q_1, q_2 with $d(p_1, p_2) = d(q_1, q_2)$ there is an isometry $f : \mathbb{H}^2 \to \mathbb{H}^2$ with $f(p_i) = q_i$.

6.3 Geodesics and hyperbolicity

As it turns out, one can describe all geodesics in $\mathbb{H}^2$, in both models.

In the half-plane model any geodesic is either a vertical segment or it is an arc contained in a half-circle orthogonal to the line $\{y = 0\}$.

In the disk model geodesics are either segments contained in diameters or arcs contained in circles orthogonal to the boundary:

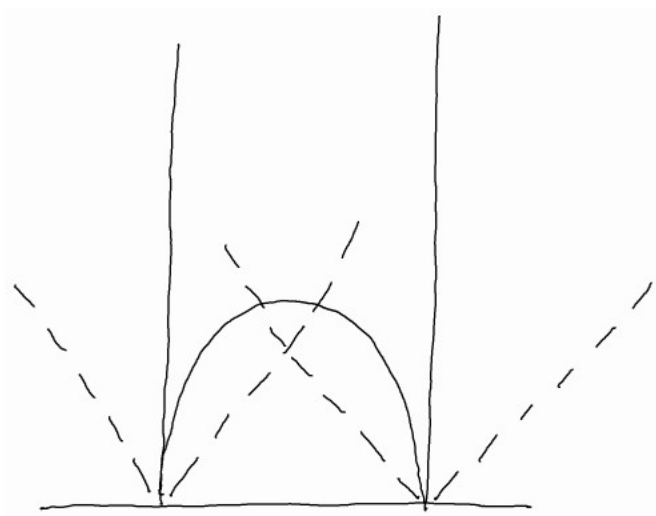


One can prove these facts without using any Riemannian geometry. Instead, one can just observe that the set of all geodesics has to be invariant under all isometries of $\mathbb{H}^2$. We will not spell this out here, as it also requires to know a it how the isometry from one model to the other one works, but I promise it's elementary.

6.3.1 Hyperbolicity

Once one has an explicit description of the geodesics, showing that $\mathbb{H}^2$ is hyperbolic becomes a reasonable task.

As a warm-up, one can consider an ideal triangle, that is to say the union, in the half-plane model, of two vertical lines and a half-circle orthogonal to $\{y = 0\}$ that looks like this:



We wish to show that any ideal triangle (there is actually only one up to isometry) satisfies the thinness condition. In order to do so, we ask ourselves how a neighborhood of one of the vertical geodesics looks like. And the answer is that it has the conical shape as in the picture. The reason is that it has to be invariant under the dilations centered at the "bottom point" of the vertical line. Increasing the angle one increases the radius of the neighborhood, and it is not hard to conclude from here finding neighborhoods as in the picture.

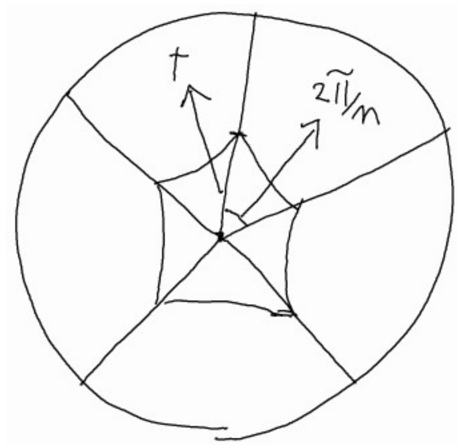
In order to deduce thinness of triangles in $\mathbb{H}^2$ from that of the ideal triangle(s), one can argue that a non-ideal triangle can be mapped isometrically inside an ideal triangle.

6.4 Right-angled n-gons and hyperbolic surfaces

Here is a funny fact about $\mathbb{H}^2$.

Lemma 6.4.1. *For every $n \geq 5$, $\mathbb{H}^2$ contains a regular right-angled n-gon.*

In order to prove it, it is convenient to use the disk model, and to consider the following family of n-gons depending on the parameter t:



(Let's say that t is the Euclidean length.) For t going to 0, the angles tend to the angles of a regular Euclidean n-gon, i.e. $(n-2)\pi/(2n)$. For t going to 1, the angles go to 0 instead. So, somewhere in between, we run into a right-angled n-gon, which is furthermore regular because of the symmetry of the picture. For this argument it is important that angles measured in the Riemannian metric coincide with angles measured in the Euclidean one, because the hyperbolic metric is a pointwise rescaling of the Euclidean one.

One can use the lemma to show that the surface called "pair of pants" admits a hyperbolic metric with totally geodesic boundery, meaning a Riemannian metric locally modelled on $\mathbb{H}^2$ or, at the boundary, on a half-plane in $\mathbb{H}^2$. A pair of pants is a 2-sphere with three open disks removed, and cutting along the lines suggested in the picture one obtains two hexagons.

One can check that gluing two regular right-angled hexagons one obtains a metric with the properties described above. The right angles play an important role in this (why?).

The reason we care about pairs of pants is that the closed (i.e. compact without boundary) surface of genus $g \geq 2$ can be decomposed into pairs of pants. Here is one way to do it for $g = 3$:

In particular, one can glue together the hyperbolic metrics on pairs of pants that we built a minute ago and show:

Theorem 6.4.2. *For any $g \geq 2$, the closed hyperbolic surface of genus g admits a hyperbolic metric, i.e. a metric locally modelled on $\mathbb{H}^2$, i.e. a metric with curvature -1 at every point.*

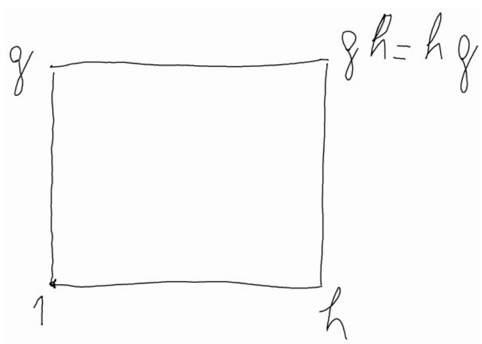
The universal cover of such a metric is going to be $\mathbb{H}^2$, and the fundamental group of the surface acts on the universal cover properly and coboundedly, whence the Milnor-Švarc Lemma and quasi-isometry invariance of (Gromov-) hyperbolicity yield:

Corollary 6.4.3. *For any $g \geq 2$, the fundamental groups of the closed hyperbolic surface of genus g is hypebolic.*

Chapter 7

Commuting stuff in hyperbolic groups

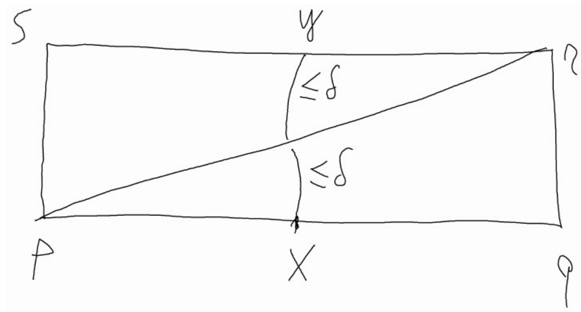
Whenever one has two commuting elements g, h in a group G, one can consider the following geodesic quadrangle in a Cayley graph:



This is a way to translate an algebraic fact, commutation, into a geometric object, namely a quadrangle. If one knows something about geodesic quadrangles in a Cayley graph, then one might try to deduce something about elements that commute in the corresponding group. We will do this for hyperbolic groups in this chapter. Notice that we have some information about geodesic quadrangles in hyperbolic spaces, namely that they are 2δ-thin:

Remark 7.0.4. Let X be δ-hyperbolic and consider a geodesic quadrangle $[p,q] \cup [q,r] \cup [r,s] \cup [s,p]$ in X. Then for any point $x \in [p,q]$ there is $y \in [q,r] \cup [r,s] \cup [s,p]$, i.e. on one of the other sides, so that $d(x,y) \leq 2\delta$. In order to show it, just split the quadrangle into two triangles using a diagonal as suggested in the picture below.

From the Remark we see that quadrangles in hyperbolic spaces are somewhat degenerate, and look very different than in, say, $\mathbb{Z}^2$. This seems to indicate that



it is very rare for elements in a hyperbolic group to commute. We will see that this is actually the case.

7.1 Results

We start with a Theorem that will be proven in the next section. It does not involve commuting elements, but we will see that it has several corollaries that do.

Theorem 7.1.1. *Let G be a hyperbolic group and suppose that $g \in G$ has infinite order. Then $\langle g \rangle$ is undistorted in G. Equivalently, the map*

$$\mathbb{Z} \to G$$
$$n \mapsto g^n$$

is a quasi-isometric embedding with respect to any given word metric d on G.

Notice that we always have $d(1, g^n) \leq |n| d(1, g)$ (if g can be written as a product of k generators then g^n can be written as a product of $|n|k$ generators), so that the content of the Theorem is that there is a lower bound on $d(1, g^n)$ which is linear in n. Oh, and we only need to worry about $d(1, g^n)$ and not more general $d(g^m, g^n)$ because $d(g^m, g^n) = d(1, g^{n-m})$.

Corollary 7.1.2. *Let G be hyperbolic and suppose that $g \in G$ has infinite order. Then if g^n is conjugate to g^m, we have either $m = n$ or $m = -n$.*

Proof. If there was some h so that $g^n = hg^m h^{-1}$, say with $|n| > |m|$, then for each positive integer k we would have:

$$g^{n^k} = hg^{m^k} h^{-1}.$$

But then we would also have, for some constants K, C not depending on k:

$$\frac{|n|^k}{K} - C \leq \left| g^{n^k} \right|_S \leq 2|h|_S + \left| g^{m^k} \right|_S \leq 2|h|_S + K|m|^k + C,$$

where S is a finite generating set for G and $|\cdot|_S$ denotes the word length (or, equivalently, the distance from 1 in $Cay(G, S)$). This is impossible for k large enough, the term on the left diverges faster than the one on the right. □

The following corollary tells us that centralisers are as small as possible.

Corollary 7.1.3. *Let G be hyperbolic and suppose that $g \in G$ has infinite order. Then the centraliser $C(g) = \{h \in G : hg = gh\}$ contains $\langle g \rangle$ as a finite index subgroup.*

Proof. We will show that there exists K so that for any $h \in C(g)$, $h\langle g \rangle$ intersects the ball of radius K around 1 in (G, d_S), for a fixed word metric d_S on G. As such ball has finitely many elements, we get that every coset of $\langle g \rangle$ in $C(g)$ has a representative in some fixed set with finitely many elements, what we wanted.

Pick any $h \in C(g)$ and choose n so that $|g^n|$ is much larger than $|h|$ (where $|\cdot|$ denotes the word length). Consider any geodesic quadrangle with vertices $1, g^n, h, g^n h = hg^n$,

Notice that the midpoint $p \in [1, g^n]$ is 2δ-close to some $q \in [h, hg^n]$, because it cannot be 2δ to $[1, g^n]$ or $[h, hg^n]$. From Theorem 7.1.1 and the fact that quasi-geodesics in hyperbolic spaces stay close to geodesics, we get that there exists D (independent of n) so that $d(p, g^i), d(q, hg^j) \leq D$ for some i, j. So, we have:

$$d(1, hg^{j-i}) = d(1, g^{-i}hg^j) = d(g^i, hg^j) \leq 2D + 2\delta,$$

so we are done setting $K = 2D + 2\delta$. (In the first equality we used that h commutes with g, and in the second one we used the left-invariance of the metric.) □

Corollary 7.1.4. *Let G be hyperbolic. Then the centre of G is virtually cyclic.*

Note that finite groups are virtually cyclic.

Proof. If G is finite, then the statement is clearly true. If G is infinite, we will prove in a later chapter that it contains some inifnite order element. The centraliser of such element contains the centre. □

We will show at some point in the future that a hyperbolic group is either finite, virtually $\mathbb{Z}$ or it contains a copy of the free group on two generators, which motivates the following corollary.

Corollary 7.1.5. *Let G be hyperbolic and suppose that G contains a copy of the free group F_2 on two generators. Then the centre of G is finite.*

Proof. Let $a, b \in G$ form a basis of $F_2 < G$. Then the centre of G is contained in the intersection of the centralisers of a, b. Using that such centralisers are virtually cyclic, it is readily checked that their intersection is finite. □

7.2 Proof of Theorem 7.1.1

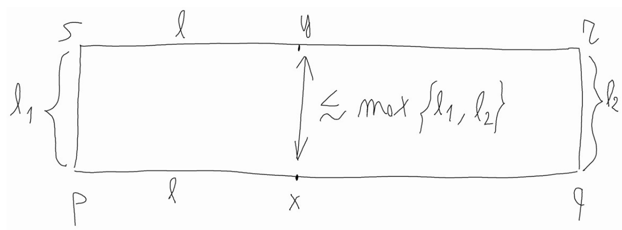
In this section we prove Theorem 7.1.1. We split the proof into three parts. We fix a hyperbolic group G and an infinite order element $g \in G$. We will work in a fixed Cayley graph of G and denote the word length by $|\cdot|$.

Part 1: $[1, g^n]$ **pretty much commutes with** g**.** We start by showing the following lemma which says that any point on a geodesic connecting points in $C(g)$ is close to some point in $C(g)$ (when this happens one says that $C(g)$ is quasiconvex).

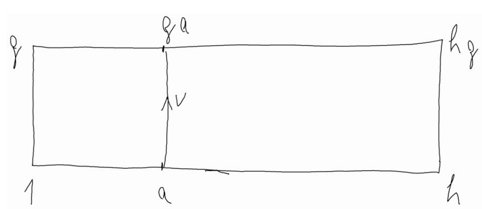
Lemma 7.2.1. *There exists D with the following property. Let $h \in C(g)$. Then $[1, h] \subseteq N_D(C(g))$.*

We will use the following fact (coarse convexity of hyperbolic metrics), that is left as an exercise:

Exercise. Consider a geodesic quadrangle $[p, q] \cup [q, r] \cup [r, s] \cup [s, p]$ in a δ-hyperbolic space. Suppose that $x \in [p, q], y \in [r, s]$ satisfy $d(p, x) = d(s, y)$. Then $d(x, y) \leq \max\{d(p, s), d(q, r)\} + 10\delta$.



Proof. Pick any $a \in [1, h] \cap G$. We claim that $d(a, ga) \leq |g| + 10\delta$. In fact, we can apply the exercise above to this geodesic quadrangle:



By definition of the distance in the Cayley graph, there exists u so that

$$ga = au$$

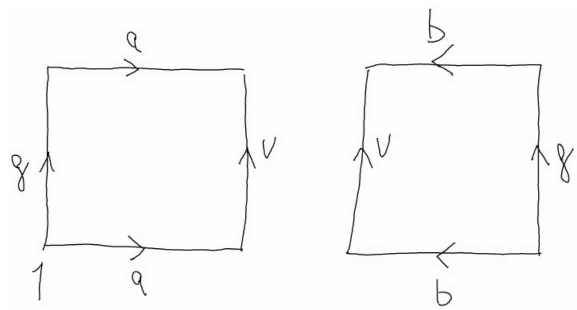
and $|u| \leq |g| + 10\delta$ (you know, distance 1 means you need to multiply on the right by one generator, distance 2 means you need two generators and so on).

Remember that we want to find some element that commutes with g and is not too far from a. So we need some way to construct elements that commute with g.

Speaking of which, here is an observation: If $gb = bu$ then ab^{-1} commutes with g.

You can check this algebraically, like so: $a(b^{-1}g) = (au)b^{-1} = gab^{-1}$, but I prefer the pictorial way: just stick together the quadrangles depicted below.

Now, pick b such that $gb = bu$ that minimises $|b|$. Our aim is now to find a bound for $|b|$. More specifically, we want to show that $|b|$ is bounded by the



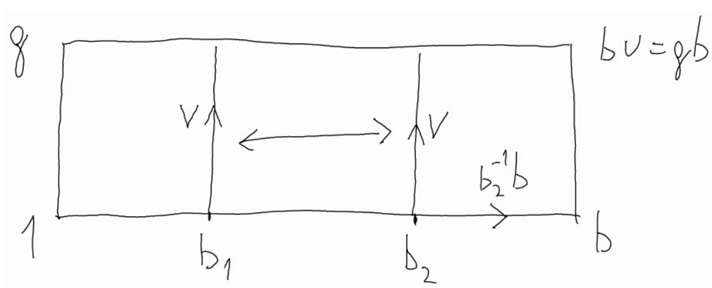
cardinality K of the ball of radius $|g| + 10\delta$ in G, which is some finite integer. In fact, for every $b_0 \in [1, b] \cap G$ there exists v in such ball so that $gb_0 = b_0 v$, again due to the exercise above. So, if $|b|$ was larger than K, there would be distinct $b_1, b_2 \in [1, b]$ and the same v in the ball so that

$$gb_i = b_i v,$$

for $i = 1, 2$. Assuming that b_1 is closer to 1 than b_2, it's not hard to check that:

- $b' = b_1(b_2^{-1}b)$ satisfies $gb' = b'u$, and
- $|b'| < |b|$.

Doing it algebraically is a bit tedious, so we justify this pictorially: Just stick together the first and third quadrangle in the picture below. □

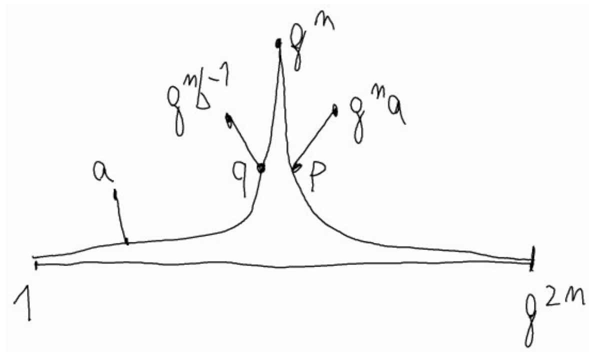


Part 2: $|g^{2n}| \approx 2|g^n|$**.**

Lemma 7.2.2. *There exists C with the followng property. For every $n \geq 0$, we have $d(g^n, [1, g^{2n}]) \leq L$.*

Proof. Set $C = 100D + 100\delta + 1$, for D as in Lemma 7.2.1. Suppose by contradiction that $d(g^n, [1, g^{2n}]) > C + \delta$. Pick $p \in g^n[1, g^n]$ so that $d(p, g^n) = C$. Notice that p cannot be δ-close to $[1, g^{2n}]$, hence it is δ-close to some $q \in [1, g^n]$. By Lemma 7.2.1, there are $a, b \in C(g)$ satisfying $d(p, g^n a), d(q, g^n b^{-1}) \leq D$.

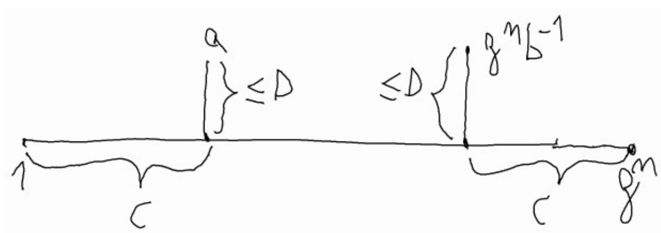
In particular,



$$|ba| = d(g^n b^{-1}, g^n a) \leq 2D + \delta.$$

Write $g^n = a(a^{-1}(g^n b^{-1}))b = arb$. From the picture below one can see that

$$|r| \leq |g^n| - 2C + 2D.$$



We also have $g^n = a^{-1}g^n a = rba$. Hence,

$$|g^n| \leq |r| + |ba| \leq (|g^n| - 2C + 2D) + (2D + \delta) < |g^n|,$$

a contradiction. □

Endgame: Information on powers of 2 suffices. Recall that our aim is to show that $|g^m|$ is comparable with m up to additive and multiplicative constants, i.e. we need constants A, B so that

$$\frac{|m|}{A} - B \leq |g^m| \leq A|m| + B.$$

The upper bound always holds, so we only need to focus on the lower bound. Also, it is enough to consider positive values of m.

A consequence of Lemma 7.2.2 is that we have, for some constant C, $|g^{2n}| \geq 2|g^n| - C$ for each n.

Let n be so that $|g^n| \geq 100C$ (which exists because g has infinite order). Set $K = \max\{|g^i| : i = 0, \ldots, n\}$.

All we are about to do now is estimating $|g^m|$ by writing m as a sum of terms of the form $2^k n$ and a remainder term, and it turns out that we can get the estimate we need in this way.

Suppose $m = \sum(2^{k_i} n) + r$, with $k_1 > k_2 > \cdots \geq 0$ and $0 \leq r \leq n$. If $k_2 < k_1 - 1$ then we have

$$\sum_{i>1} 2^{k_i} \leq \sum_{j<k_1-1} 2^j \leq 2^{k_1-1}$$

and the estimate is direct:

$$\begin{aligned}|g^m| \geq 2^{k_1}|g^n| - k_1 C - \sum_{i>1} 2^i |g^n| - K \geq \\ 2^{k_1-1}|g^n| - k_1 C - K \geq \\ \frac{|g^n|}{10n} m - K.\end{aligned}$$

The last inequality comes from $2^{k_1-1} \geq m/(4n)$ and $|g^n|$ being much larger than C.

If $k_2 = k_1 - 1$, then we re-write $m = 2^{k_1+1}n - 2^{k_2}n + \sum_{i\geq 3}(2^{k_i}n) + r$, and proceed similarly. □

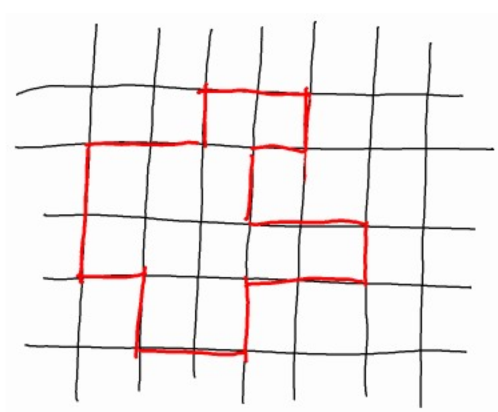
Chapter 8

Geometry of presentations

The main idea we will exploit in this chapter is that whenever you have a group G generated by the symmetric set S and a word w in the alphabet S that represents the identity in G, then w gives a loop in $Cay(G,S)$. Actually there's a $1-1$ correspondence between such words and combinatorial loops in $Cay(G,S)$.

Notice that in the previous chapter we exploited the special case of this principle when the word represents the commutator $[g,h]$ of two commuting elements g,h.

In order to motivate what follows, let us draw a loop in the Cayley graph of $\mathbb{Z}^2$ with respect to standard generators that we denote by a,b:



One is almost tempted to say that the loop is homotopically trivial, except that the squares you see are not actually there. What's true is that if you fill in all squares, then the loop we drew, as well as any other loop, becomes trivial. Any combinatorial loop in a Cayley graph has a natural label associated to it, defined just by writing down the generators corresponding to the edges traversed by the loop. Notice that the label of the boundary of the squares in the picture above is (a cyclic permutation of) $[a,b] = aba^{-1}b^{-1}$. Oh, wait, but $\mathbb{Z}^2 = \langle a,b|[a,b]\rangle$...

Let us try to generalise.

Definition 8.0.3. Let $\langle S|R\rangle$ be a group presentation. Its *Cayley complex* $Cay_c(\langle S|R\rangle)$ is obtained from the Cayley graph by gluing disks, that we call *tiles*, to all loops labelled by some $r \in R$.

And the phenomenon we observed above has the following generalisation:

Proposition 8.0.4. *Any Cayley complex is simply connected. Conversely, given a group with generating set S and a set of words on S representing 1 in G, gluing disks to all loops labelled by some $r \in R'$ makes $Cay(G,S)$ simply connected if and only if the kernel of the natural map $F_S \to G$ is normally generated by R'.*

The second part is a statement that will be needed later.

We will not prove the proposition, but I'd still like to explain the reason why it's true: The fundamental group of $Cay(G,S)$ is isomorphic to $K = Ker(F_S \to G)$. The map from K to $\pi_1(Cay(G,S))$ can be descried as follows. An element of K corresponds to a word in S that represents the identity in G, whence a loop in the Cayley graph based at the identity.

Once one accepts this fact, the first part of the proposition can be proven analysing the effect on $\pi_1(Cay(G,S)) = \langle grg^{-1}\rangle_{g\in G, r\in R}$ of gluing a disk based at some $g \in G$ and with label r. And such effect is just "killing" one of the generators. The second part is similar.

8.1 The algebraic point of view

Let's fix, for the purposes of this section, a presentation $\langle S|R\rangle$. If we have a word w in the alphabet $S \cup S^{-1}$ that represents the identity, then by definition we can write it as a product of conjugates of relators:

$$w =_{F_S} \prod_{i=1}^{k} g_i r_i g_i^{-1}, \quad (*)$$

for some $g_i \in G$ and $r_i \in R$. Let me emphasise that the equality holds in the free group on S. We just saw that w corresponds to a loop in $Cay(G,S)$ that "spans" a disk in $Cay_c(S,R)$, as the latter is simply connected.

So, are these two facts related? Yes, indeed! Let me try to informally explain this.

Suppose we have $(*)$, and let us construct a disk whose boundary is the loop l in $Cay(G,S)$ (based at 1 and) whose label is w. First of all, let w' be the word on the right-hand side, and let l' be the corresponding loop in $Cay(G,S)$. We know that it is obtained from w using free reductions/insertions (meaning substitutions like $uss^{-1}v \to uv$, where $s \in S$, or the inverse procedure $uv \to uss^{-1}v$). You should convince yourself, using induction, that this makes

l and l' homotopic, meaning that there is an annulus whose outer circle maps to l and inner circle maps to l'. So, we now wish to fill w' instead of w. This is possible because the loop corresponding to each term $g_i r_i g_i^{-1}$ looks like Figure 8.1.

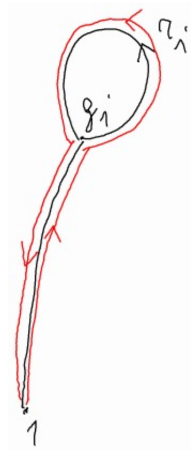


Figure 8.1

By the definition of the Cayley complex, there is a disk filling the loop in the picture. Hence, l' can be filled by a disk whose image, schematically, looks like Figure 8.2.

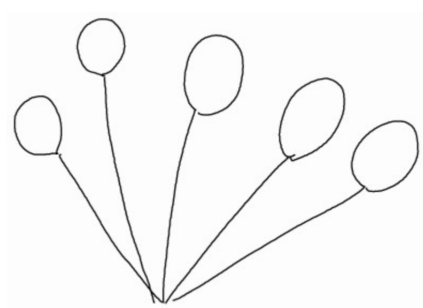


Figure 8.2

The picture is oversimplified, because the loops and sisk can overlap, and will in general.

This was a brief explanation about how from $(*)$ one gets a map from a disk in $Cay_c(S,R)$. Let us now consider a "nice" disk in $Cay_c(S,R)$, meaning one whose image is naturally tiled by the tiles of $Cay_c(S,R)$, as in Figure 8.3:

Let us consider an "external tile", as in the picture. The important thing to notice is that the labels of the loops l_1 and l_2 differ by multiplication by grg^{-1} and free reductions. The loop labelled grg^{-1} is the green one in the picture, it

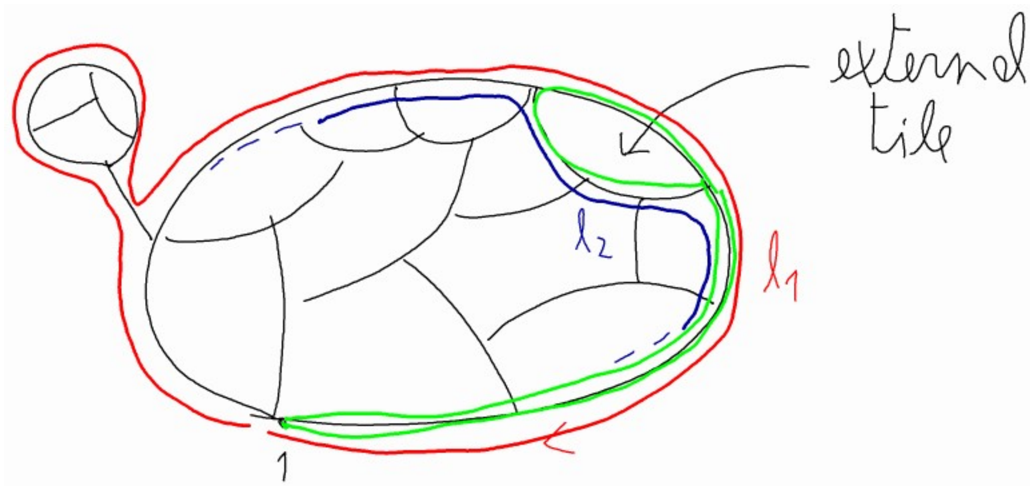


Figure 8.3

travels back along l_1, loops around the external tile and then goes back to 1.

What we just described is excatly the inductive step needed to write the label of the boundary of a disk as a product of conjugates of relators.

8.2 Finite presentations

Suppose that we have a group G given by a *finite presentation* $\langle S|R\rangle$, meaning that S, R are required to be finite. Then Proposition 8.0.4 seems to suggest that $Cay(G,S)$ is simply-connected "up to bounded holes". There's a way to make this precise, as we're about to see. We would like to say that that any loop spans a "disk with bounded holes", and we are going to call such objects k-coarse disks.

Definition 8.2.1. A map $f : D^2 \to X$ from the unit disk in $\mathbb{R}^2$ into the metric space X is a *k-coarse disk* if there exists $\epsilon > 0$ so that whenever $x.y \in D^2$ satisfy $d(x,y) \leq \epsilon$, we have $d(f(x), f(y)) \leq k$.

The geodesic metric space X is *k-coarsely simply connected* if for every loop $l : S^1 \to X$ there exists a k-coarse disk $f : D^2 \to X$ that fills it, meaning that $f|_{\partial D^2 = S^1} = l$.

Exercise: Being coarsely simply connected is a quasi-isometry invariant for geodesic metric spaces.

Recall that a group is finitely presented if it admits a finite presentation, that is to say it is isomorphic to $\langle S|R\rangle$ for some finite S, R. Here is the main result of this section.

Theorem 8.2.2. *The group G is finitely presented $\iff$ one of its Cayley graphs is coarsely simply connected $\iff$ all its Cayley graphs are coarsely simply connected.*

The second " $\iff$ " is a consequence of the exercise above, which also gives us that:

Corollary 8.2.3. *Being finitely presented is a quasi-isometry invariant.*

That's nice, we have an algebraic property that turns out to be a quasi-isometry invariant...

Let us prove the first " $\iff$ " of the theorem.

Proof. $\Rightarrow$: Suppose that G is (isomorphic to) the group given by the finite presentation $\langle S|R\rangle$. We know from Proposition 8.0.4 that the Cayley complex $Cay_c(S,R)$ is simply connected. Consider now a loop $l : S^1 \to Cay(G,S)$ (and recall that we have to show that we can fill it by a coarse disk). There exists a continuous map $f : D^2 \to Cay_c(S,R)$ that restricts on ∂D^2 to l. We can now construct a new map $\hat{f} : D^2 \to Cay(G,S)$ in such a way that if $f(x)$ lies inside a tile, then $\hat{f}(x)$ lies on the boundary of the same tile.

It is easy to see that $\hat{f}$ is a k-coarse disk, where k only depends on the maximal length of a relator in R, which is finite because the presentation is finite.

$\Leftarrow$: Suppose that $Cay(G,S)$ is k-coarsely simply connected. Let R be the set of labels of all combinatorial simple loops in $Cay(G,S)$ of length at most $10k+10$. We claim that $\langle S|R\rangle$ is a presentation for G. By the second part of Proposition 8.0.4, what we need to show is that when we glue disks to all simple loops in $Cay(G,S)$ as above, we get a simply connected space X.

It is enough to check that combinatorial loops in X span disks, because the fundamental group of X is generated by (elements represented by) combinatorial loops.

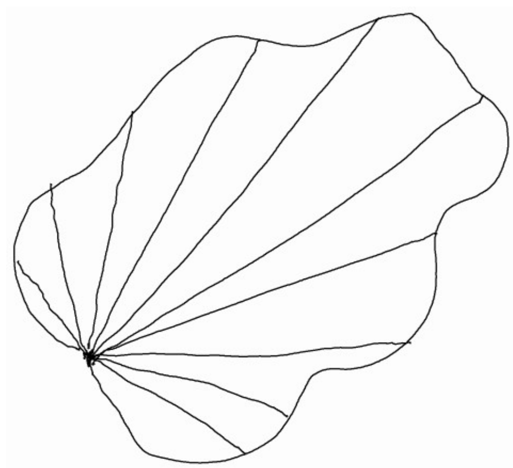
Let l be a combinatorial loop in $Cay(G,S)$ and let $f : D^2 \to Cay(G,S)$ be a k-coarse disk that fills it. We have to turn this coarse disk in $Cay(G,S)$ into a continuous disk g in $Cay_c(S,R)$, by "filling in the holes". Let us triangulate D^2 in a sufficiently fine way, and let us perturb f to assume that the image of vertices in such triangulation are vertices of $Cay(G,S)$. Let us now define a continuous map $g : D^2 \to Cay_c(S,R)$. First of all, g coincides with f on ∂D^2 and on the vertices of the triangulation. On each edge contained in the interior of D^2, g is a geodesic. To show that the map we defined can be extended to the whole of D^2 we now only need to convince ourselves that the loop obtained restricting g to the boundary of a triangle of the triangulation is homotopically trivial in $Cay_c(S,R)$. This follows from easily checked fact that the length of such loop is at most $10k+10$ (if the triangulation we started with was fine enough). We now that *simple* loops of length bounded by $10k+10$ are homotopically trivial in $Cay_c(S,R)$, but it's not hard to see that this implies that all loops with the same length bound are homotopically trivial as well. □

8.3 Finite presentations of hyperbolic groups and Dehn functions

Let us now get back to hyperbolic groups. Now that we have Theorem 8.2.2, it's not that hard to show:

Theorem 8.3.1. *Hyperbolic groups are finitely presented.*

Without spelling out the details, here is probably the simplest possible proof of the fact that the Cayley graph of a hyperbolic group is coarsely simply connected. Take any loop, and construct a coarse disk using geodesics connecting the basepoint of the loop to points along the loop as suggested in the following picture:



This works because geodesics connecting nearby points in a hyperbolic space stay close to each other. This is a very mild consequence of hyperbolicity, actually. And in fact, on one hand this proof applies to more general groups, and on the other it is not optimal for hyperbolic groups in a sense that we are about to specify.

We would like now to quantify how hard it is to fill loops in the Cayley complex.

Definition 8.3.2. Let $\langle S|R\rangle$ be a finite presentation. Define the area $Area(w)$ of a word w in $S \cup S^{-1}$ that represents 1 in $\langle S|R\rangle$ to be the minimal k so that we can write $w =_{F_S} \prod_{i=1}^{k} g_i r_i g_i^{-1}$ for some $g_i \in G, r_i \in R$. This coincides with the minimal number of tiles contained in the image of a disk in $Cay_c(S,R)$ with boundary label w. Finally, define the *Dehn function* as

$$\delta_{S,R}(n) = \max_{w:\|w\|_S \leq n} Area(w),$$

where $\|w\|_S$ is length of the word w.

You should try to convince yourself that the Dehn function of the standard presentation of $\mathbb{Z}^2$ is quadratic.

The Dehn function depends on the finite presentation, not just the group being presented. As it turns out, however, a properly defined "asymptotic class" of the Dehn function depends on the group only, not the specific presentation (as long as the said presentation is finite). We will not go into that, let me just mention that it makes sense to say that a group has, for example, linear, quadratic or exponential Dehn function.

Back to hyperbolic groups. The proof sketched above can be made quantitative, and it gives a quadratic Dehn function. However, the Dehn function is much smaller than that, and there's even more:

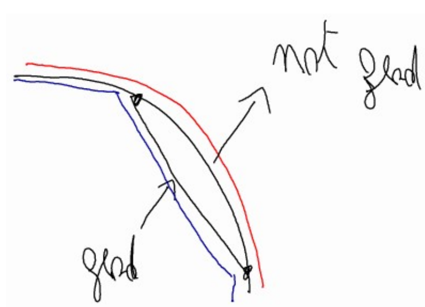
Theorem 8.3.3. *Let $\langle S|R\rangle$ be a finite presentation of a hyperbolic group. Then $\delta_{S,R}$ is (at most) linear. Moreover, if a group admits a finite presentation with linear Dehn function, then it it hyperbolic.*

So, if a word represents the identity in a hyperbolic group, then there is a very efficient way to actually write it down as a product of conjugates of relators. Also, the "moreover" part gives a convenient criterion to check that a group is hyperbolic, and we will exploit it in the next chapter.

We will not prove the moreover part. The remainder of this chapter is devoted to the outline of the bound on the Dehn function of a hyperbolic group.

We will say that a function $\gamma : I \to X$, where I is a closed interval in $\mathbb{R}$ and X is a metric space, is a K-*local geodesic* if whenever $|x-y| \leq K$ we have $d(\gamma(x), \gamma(y)) = |x-y|$.

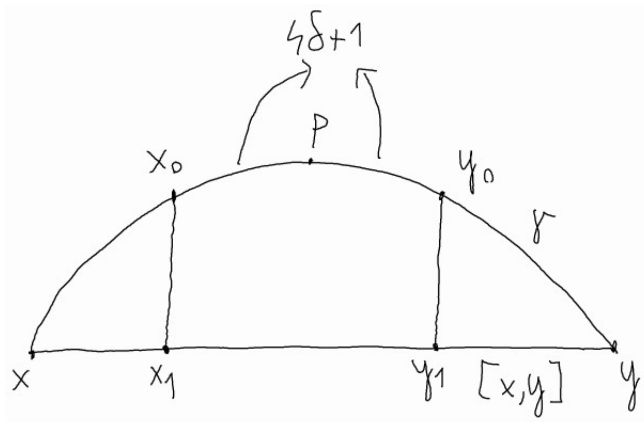
The reason we care is that if we have a combinatorial loop in a Cayley graph which is NOT a K-local geodesic, then we can reduce its length by at least 1 by a "local move" as suggested in the picture:



Also, there is some N depending only on K and the presentation so that there is a homotopy between the two loops in the picture involving at most N tiles of the Cayley complex. Hence, you see that if we knew that there exists some K some that any loop is not a K-local geodesic, then we could iterate this procedure and fill any given loop with a number of disks that is linear in the length of the loop. It is almost obvious from the following lemma that there exists such K.

Lemma 8.3.4. *Let X be a δ-hyperbolic space. Then any $(8\delta+2)$-local geodesic γ in X lies within distance $4\delta+1$ from any geodesic with the same endpoints.*

Proof. Let x, y be the endpoints of γ, and let $[x, y]$ be any geodesic connecting them. Let us pick the "worst point" $p \in \gamma$, meaning the one that is further away from $[x, y]$. If $d(p, x) \leq 4\delta + 1$ or $d(p, y) \leq 4\delta + 1$, then we are done. Hence, suppose that this is not the case and pick points $x_0, y_0 \in \gamma$ as in the picture, letting $x_1, y_1 \in [x, y]$ be points minimising the distance from x_0, y_0.



There is a geodesic quadrangle in the picture, and p is on one of its sides. So, it is 2δ-close to one of the three other sides, and we would like to argue that it cannot be 2δ-close to either $[x_0, x_1]$ or $[y_0, y_1]$. In particular, it is going to be 2δ-close to $[x, y]$, as we wanted.

Suppose by contradiction that there is a point q on, say, $[x_0, x_1]$, so that $d(p, q) \leq 2\delta$. Notice that $d(x_0, q) \geq 2\delta + 1$, and hence

$$d(x_1, q) \leq d(x_0, x_1) - 2\delta - 1 \leq d(p, [x, y]) - 2\delta - 1.$$

Oops, but then

$$d(p, [x, y]) \leq d(x_1, p) \leq d(x_1, q) + d(q, p) < d(p, \gamma),$$

a contradiction. □

Let me conclude this chapter by saying that the outline above also works to show another theorem. The word problem (for a given presentation) asks whether there exists an algorithm to determine whether a given word represents the identity or not. Surprisingly, there are rather explicit presentations that one can write down where the problem is not solvable! But for hyperbolic groups...

Theorem 8.3.5. *The word problem in a hyperbolic group is solvable in linear time.*

... not only the word problem is solvable, but it can be solved very efficiently.

Can you see how one can prove the theorem relying on the property of local geodesics in hyperbolic spaces?

Chapter 9

Small cancellation

In this chapter we construct examples of hyperbolic groups. As it turns out, there is an easy-to-check and widely applicable criterion for a presentation to give a hyperbolic group.

Roughly speaking, the criterion is that common subwords of distinct relators should be short compared to the length of the relators themselves. We will give the formal definition after we setup some language.

Let S be a finite set, and denote by $W(S)$ the set of all words in the alphabet $S \cup S^{-1}$. A subset $R \subseteq W(S)$ is said to be *symmetrised* if

- each $w \in R$ is a reduced word,
- if $w \in R$, then w^{-1} is in R as well,
- if $w \in R$, then all its cyclic shifts are in R as well.

The last item means that if w can be written as avb, then bav is also in R (in particular, it is reduced). Notice that the elements of the free group on S corresponding to avb and bav are conjugate, as $avb =_{F_S} b^{-1}(bav)b$.

Definition 9.0.6. Fix some $R \subseteq W(S)$. A *piece* is a word $w \in W(S)$ that appears as an initial subword of two distinct words in R.

For example, if $R = \{a^2b, a^3\}$, then a and a^2 are pieces, but a^3 is not.

Finally, here is the key definition.

Definition 9.0.7. A symmetrised collection of words $R \subseteq W(S)$ satisfies $C'(\lambda)$, where $\lambda > 0$, if for each $r \in R$ and each piece w that appears as a subword of r we have $\|w\|_S < \lambda \|r\|_S$.

The picture you may wish to keep in mind is that relators give loops in the Cayley graph and when two such loops intersect we have:

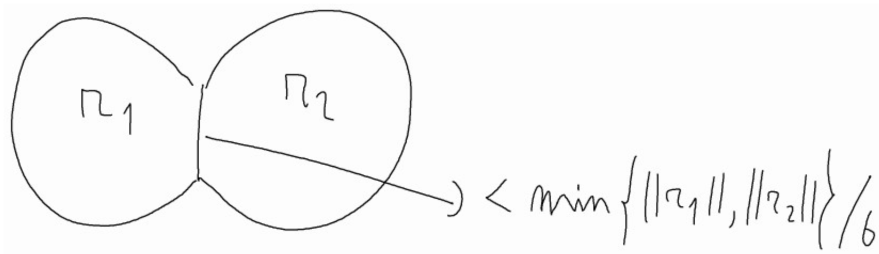


Figure 9.1

And even more finally, here is the theorem.

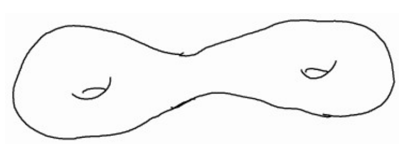
Theorem 9.0.8. *Suppose that $R \subseteq W(S)$ satisfies $C'(1/6)$ and is finite. Then the group $\langle S|R\rangle$ is hyperbolic.*

The plan for the rest of this chapter is to first give examples of presentations satisfying $C'(1/6)$ and then to outline the proof of the theorem.

9.1 Examples 1,2,...

Suppose that $R \subseteq W(S)$ is so that every $r \in R$ is cyclically reduced, meaning that r and all its cyclic shifts are cyclically reduced. In this case we denote by $Symm(R)$ the smallest symmetrised subset of $W(S)$ containing R. Notice that $\langle S|R\rangle$ is the same group as $\langle S|Symm(R)\rangle$ (recall that cyclic shifts correspond to conjugates in F_S).

Here is our first example of $C'(1/6)$ collection of words. Let $S = \{a, b, c, d\}$, and let $R = \{[a, b][c, d]\}$. Then $Symm(R)$ satisfies $C'(1/6)$. In fact, all words in $Symm(R)$ have length 8, and the maximal length of a piece is readily checked to be 1. So, $G = \langle a, b, c, d|[a, b][c, d]\rangle$ is hyperbolic. Do you recognise this group? It's the fundamental group of the closed (oriented, connected) surface of genus 2.



We knew already that G is hyperbolic, but this is an entirely different proof... Fundamental groups of higher genus surfaces also have presentations satisfying $C'(1/6)$.

Let us move on to the next example. Fix some alphabet S (of finite cardinality at least 2) and denote by $W_{k,l}(S)$ the collection of all k-tuples of cyclically reduced words of length l in $S \cup S^{-1}$. Then, as it turns out,

$$\frac{\#\{R \in W_{k,l}(S) | Symm(R) \text{ satisfies } C'(1/6)\}}{\#W_{k,l}(S)} \xrightarrow{l \to +\infty} 1. \quad (*)$$

In particular, *"almost every" group admitting a presentation with $\#S$ generators and k relators is hyperbolic!* Cool, right?

It is not hard to prove $(*)$, it's just a counting argument (you're welcome to try!). The intuition for why it's true is that if you write down two words randomly (meaning that you always choose the next letter using a dice or something), then it is unlikely that they contain long common subwords.

We will give a third example later.

9.2 Proof of Theorem 9.0.8

I managed to avoid it so far, but I'm afraid that it's time to define disk diagrams. We informally used them in the previous chapter in the form of "nice" disks in a Cayley complex.

Let $\langle S|R\rangle$ be a presentation, fixed from now on, where R is symmetrised. A *disk diagram* is a contractible planar 2-complex D where *edges* (i.e. 1-cells) are labelled by elements of S, and so that the label of the boundary of each *face* (i.e. 2-cells) is an element of R. The *boundary label* is the (cyclic) word that can be read going around the boundary of the complementary region of D in $\mathbb{R}^2$.

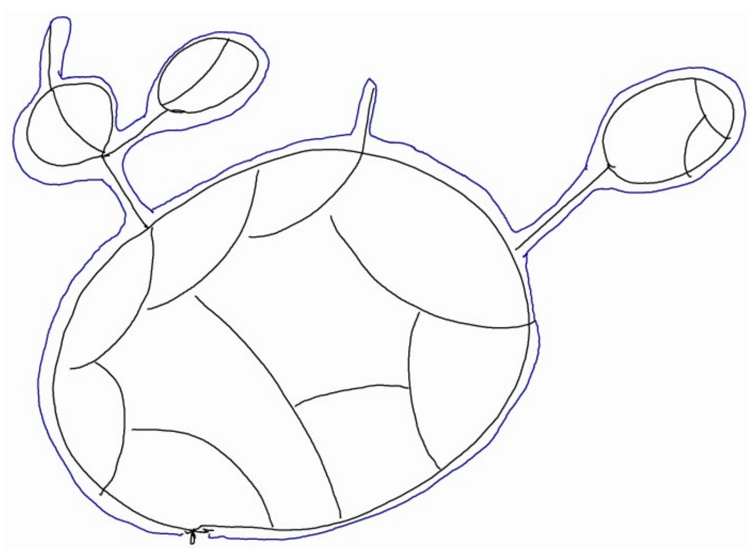


Figure 9.2: A "generic" disk diagram. The boundary label is read along the blue path.

A more combinatorial version of Proposition 8.0.4 is that, given any word representing the trivial element, there is a disk diagram whose boundary is the given word. This is known as van Kampen Lemma.

Ok, now that we defined it, you can forget the definition and pretend that disk diagrams are actual disks subdivided into 2-cells, rather than contractible, planar, etc.

The key step in the proof of Theorem 9.0.8 is the following result.

Theorem 9.2.1. *[Greendlinger's Lemma] Suppose that the (non-necessarily finite) presentation* $\langle S|R\rangle$ *satisfies* $C'(1/6)$. *Suppose that the disk diagram* D *has the minimal number of faces among all diagrams with a given boundary label, and that such number is positive. Then there exists a face* F *in* D *so that* ∂F *and* ∂D *contain a common arc of length* $> (1-3\lambda)l(\partial F)$.

(Here $l(\partial F)$ is the length of the boundary of F.)

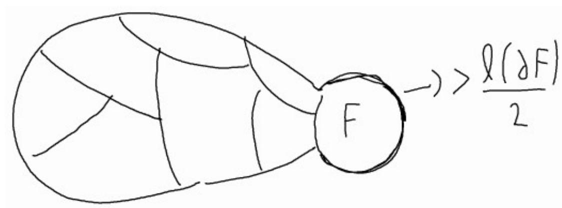
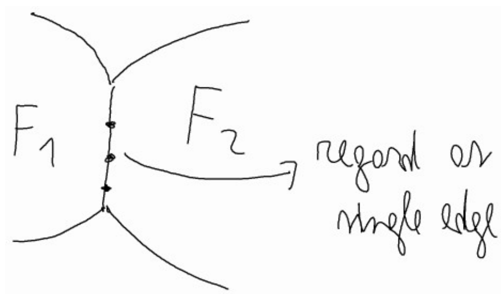


Figure 9.3: Conclusion of Greendlinger's Lemma.

Once we know Greendlinger's Lemma, it is easy to check that the Dehn function of a finite $C'(1/6)$ presentation is linear, whence that such presentations give hyperbolic groups by Theorem 8.3.3. In fact, given a (combinatorial) loop l, we can consider a minimal disk diagram D whose boundary label is the same as the label of our loop. The point of "1/6" is then that by Greendlinger's Lemma, there exists a face F in D so that more than half of its boundary is on ∂D. We can then "push l across" such face and get a strictly shorter loop. But such new loop admits a minimal disk diagram D' which is contained in D (convince yourself that it actually is minimal...). Proceeding inductively, we see that the number of faces in D is linear in the length of l, as required.

We are left to prove Greendlinger's Lemma.

"Proof" of Greendlinger's Lemma. We will make the simplifying assumption that D is topologically a disk, and we set $\lambda = 1/6$. The proof of the general case is similar but more technical. Let us now "remove edges" of valence 2, meaning that we regard intersections of faces as in the picture as a single edge.



Let us denote by v the number of vertices, e the number of edges and f the number of faces of D.

Then, just because D is homeomorphic to a disk and hence has Euler characteristic 1, we have $v - e + f = 1$.[1]

Let us start with the warm-up observation that the contribution of internal faces to $\chi = v - e + f$ is negative. Let me explain what I mean. An internal face F is a face that doesn't intersect the boundary of D. We can split the boundary of F into arcs contained in the intersection of F with some other face. The point is that each such arc has length *strictly* less than 1/6 of the length of ∂F, see Figure 9.1.[2] In particular, there are $k \geq 7$ such arcs. You can think that each edge contributes 1/2 to χ, as it is shared by two faces, while each vertex contributes at most $-1/3$, as it is shared by at least 3 faces. Hence, the total contribution to χ due to the face we're looking at is at most

$$k/3 - k/2 + 1 = 1 - k/6 < 0.$$

But χ is positive in the end. So, if there is an internal face, some other face has to compensate for the negative contribution. What we are about to do now is studying the sign of the contribution of each face. The upshot is that the positive contribution comes from the faces as in the conclusion of Greendlinger's Lemma.

[1] If you don't know what Euler characteristic is, to justify why the formula holds notice that it holds for the simplest disk diagram (one face with one edge going around it once, and just one vertex), and the formula is "stable" under taking subdivisions.

[2] There is a subtle but important point here: It might be the case that the two adjacent faces correspond to the same relator. And this means that if you fix a point in their intersection and read the boundary labels of the two faces starting from there, the boundary labels agree. But in this case, you could form a new diagram with the same boundary label by removing the two faces and identifying the two arcs with the same label that you get, contradicting minimality. (I cheated a bit.)

Suppose that we consider a face with $j \geq 1$ arcs on the boundary. Then it will also have j arcs in the interior, each of them split into, say, k_i edges. Figure 9.4 summarises the contributions to χ of all the objects involved.

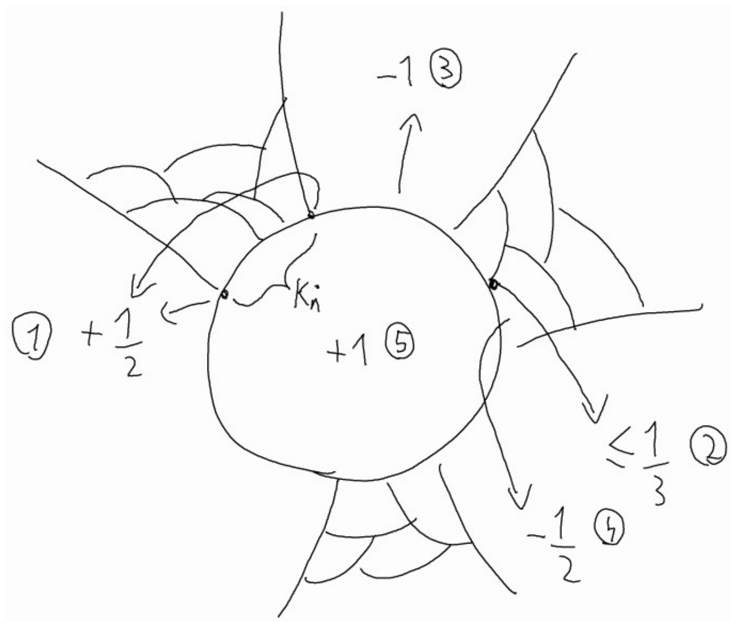


Figure 9.4: The five terms of $(*)$.

The total contribution of the face is hence

$$\leq j + \frac{\sum k_i - 1}{3} - j - \frac{\sum k_i}{2} + 1 = \frac{-\sum k_i}{6} - j/3 + 1. \quad (*)$$

In particular, for $j \geq 2$ it is easy to see that the contribution is nonpositive. Also, for $j = 1$ the formula simplifies to $-k_1/6 + 2/3$, hence the contribution is positive only if $k_1 \leq 3$.

Ok, now, χ is positive, so someone must be contributing positively to it. And as we just saw, the only option for that is a face with one arc on the boundary that intersects at most 3 other faces. Each intersection with some other face has length less than $l(\partial F)/6$, which implies that the arc on the boundary has length larger than $l(\partial F)/2$, exactly what we wanted. □

9.3 ... and 3

Here is the final example of $C'(1/6)$ presentation, due to Pride. We are about to show that there exists an infinite group G whose only finite quotient is the trivial one, a rather exotic beast. Most likely all the infinite groups you are thinking about now have many finite quotients...

Let u_i, v_i, for $i \geq 1$, be words in a^i, b^i. Let G be given by the presentation $\langle a, b | \{au_i, bv_i\}_{i \geq 1} \rangle$. We claim that G has no non-trivial finite quotients. In fact, let $\phi : G \to F$ be a surjective homomorphism onto a finite group. Then there exists n (e.g. $n = |F|$) so that $\phi(a^n) = \phi(b^n) = 1$. But then $\phi(u_n) = \phi(v_n) = 1$, which in turn implies that $\phi(a) = \phi(b) = 1$. So, the image of G, which coincides with F, must be trivial.

Ok, but now... Is G infinite? Maybe we just wrote down a very complicated presentation of the trivial group, who knows. Let me remark that there does not exist an algorithm to determine whether a presentation gives the trivial group or not (!). So, even though it may seem to you that you can probably just sit down, work a bit and show that for some choice of u_i, v_i one gets an infinite group, this can be trickier than one expects...

But there is a way to guarantee that G is infinite. Suppose that we choose words u_i, v_i in such a way that $Symm(\{au_i, bv_i\})$ satisfies $C'(1/6)$, and so that, for each k, a^k is never more than half of a relator (such choice is possible). Then an application of Greendlinger's Lemma gives that $a^n \neq_G 1$ for each $n \geq 1$, and in particular G is infinite.

Let me conclude this chapter by saying that by studying the $C'(1/6)$ condition we scratched the surface of the so-called small cancellation theory. Small cancellation theory has been used extensively to construct exotic groups, and also to construct interesting quotients of given groups.

Chapter 10

Free subgroups

Part III

CAT(0) spaces and cube complexes

Chapter 11

CAT(0) geometry

In this chapter we introduce CAT(0) and nonpositively curved spaces. Nonpositively curved spaces are meant to generalise manifolds of nonpositive sectional curvature, and CAT(0) spaces their universal covers.

The phenomenon that the definitions capture is that geodesic triangles in nonpositive curvature are not "fatter" than geodesic triangles in $\mathbb{R}^2$. This holds locally in nonpositively curved spaces and globally in CAT(0) spaces.

Another good way to think about curvature is that negative curvature forces geodesic to diverge rather fast, zero curvature describes the familiar Euclidean space, while in positive curvature geodesic start off diverging slowly and end up actually converging back together after a while (think of S^2, where geodesics emanating from the north pole are subarcs of longitudes, and hence, when prolonged, they all end up at the south pole).

11.1 The CAT(0) inequality

Given a triangle Δ in a metric space, its *comparison triangle* in $\mathbb{R}^2$, denoted $\overline{\Delta}$ is a triangle with the same side lengths as Δ.

Notice that all such triangles are isometric, so we will often talk about *the* comparison triangle, with a slight abuse. Also, we will always implictly fix a map $\phi : \Delta \to \overline{\Delta}$ which is isometric on the sides of Δ, but to keep the notation compact, instead of $\phi(p)$ we write $\overline{p}$. As Figure 11.1 is meant to illustrate, for p a point on Δ (say not a vertex), $\overline{p}$ is the point on $\overline{\Delta}$ so that

- $\overline{p}$ is on the side $\overline{\gamma}$ of $\overline{\Delta}$ that corresponds to the side γ of Δ where p lies, and
- the distances from $\overline{p}$ to the endpoints of $\overline{\gamma}$ are the same as the distances of p from the endpoints of γ.

Definition 11.1.1. A geodesic metric space X is CAT(0) if the following holds. For any geodesic triangle Δ and $p, q \in \Delta$, we have

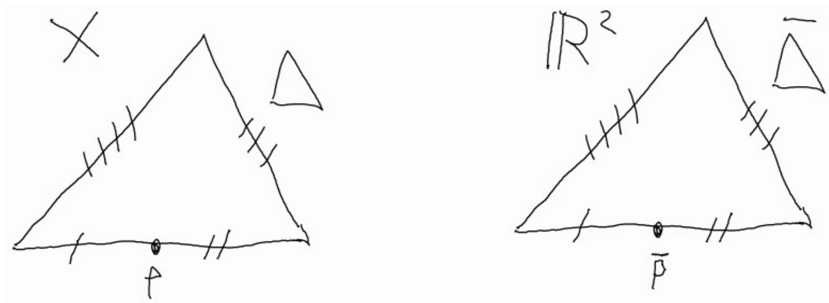
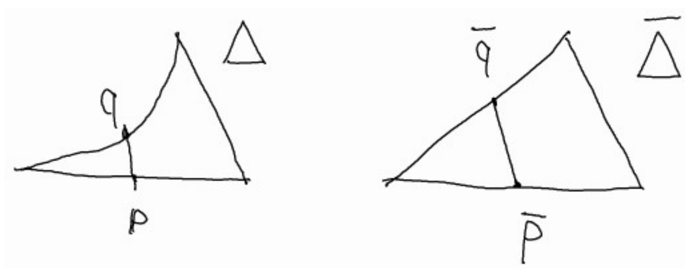


Figure 11.1

$$d_X(p,q) \leq d_{\mathbb{R}^2}(\overline{p},\overline{q}). \quad (*)$$

A *nonpositively curved* space is a geodesic metric space which is locally CAT(0) space, i.e. a space where each point has a neighborhood so that inequality $(*)$ holds for triangles contained in that neighborhood.



There is a very useful connection between CAT(0) and nonpositively curved spaces, via universal covers:

Theorem 11.1.2. *If X is nonpositively curved and simply connected, then it is* CAT(0)*. In particular, the universal cover of a nonpositively curved space is* CAT(0)*.*

11.2 First examples

The following spaces are CAT(0). We will see more examples (of a more combinatorial nature) later on.

- $\mathbb{R}^n$. That's clear, isn't it?
- $\mathbb{H}^n$. This is less obvious, but still true. And it shouldn't be too hard to believe, I hope...

- More generally, simply connected complete Riemannian manifolds where all the sectional curvatures are nonpositive.

- The following result gives a way to combine CAT(0) spaces into larger ones. We say that a subspace C of geodesic metric space X is convex if for every pair of points p, q in C the geodesics in X connecting p to q are contained in C.

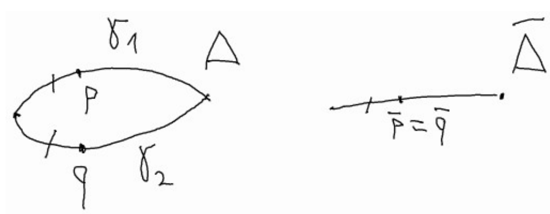
  Theorem 11.2.1. *Let X, Y be* CAT(0) *spaces and suppose that they have a common convex subspace Z. Then the union of X and Y along Z (endowed with the induced path metric) is a* CAT(0) *space.*

  You can take for example $X = \mathbb{H}^2$, $Y = \mathbb{R}^2$ and identify a bi-infinite geodesic in X with a bi-infinite geodesic in Y.

11.3 Some properties

We now show some properties of CAT(0) spaces so we get to see how one can use comparison triangles.

Lemma 11.3.1. CAT(0) *spaces are uniquely geodesic, that is to say there exists a unique geodesic connecting any two given point in a* CAT(0) *space.*



Proof. Consider a geodesic bigon in a CAT(0) space, i.e. the union of two geodesics γ_1, γ_2 with common endpoints. You can also regard such bigon as a geodesic triangle Δ with one side of length 0. But then the comparison triangle $\overline{\Delta}$ is degenerate. So, if $p \in \gamma_1, q \in \gamma_2$ are at the same distance from the common starting point of γ_1, γ_2, then $\overline{p} = \overline{q}$, which implies $p = q$, and we are done. □

Let's say that $\gamma : I \to X$, where I is a closed interval in $\mathbb{R}$ and X is a metric space, is a *local geodesic* if each t in I is contained in a non-trivial interval J so that

- $\gamma|_J$ is an isometric embedding,
- t is contained in the interior of J if it is not an endpoint of I.

Lemma 11.3.2. *Local geodesics in a* CAT(0) *space are geodesics.*

Local-to-global properties are always very useful, because local properties tend to be much easier to check.

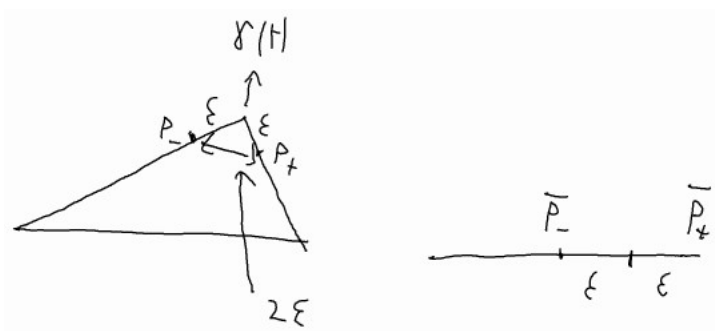
Proof. Let t_0 be the first point of I. Let $S = \{t : \gamma|_{[t_0,t]}$ is an isometric embedding$\}$. We will show that S is non-empty, closed and open, which implies that it coincides with I.

Well, actually, it is non-empty because $t_0 \in S$ and the fact that it's closed is an easy consequence of the continuity of the distance, we don't need the CAT(0) inequality for that.

Let us then show that it's open. Let $t \in S$, and we can suppose $t \neq t_0$ for otherwise we can directly use the definition of local geodesic. Pick $\epsilon > 0$ small enough that $\gamma|_{[t-\epsilon,t+\epsilon]}$ is an isometric embedding. Consider the geodesic triangle Δ with vertices $\gamma(t_0), \gamma(t), \gamma(t+\epsilon) = p_+$. We want to show that Δ is degenerate, so that we get that $\gamma|_{[t_0,t+\epsilon]}$ coincides with the geodesic from $\gamma(t_0)$ to p_+.

How do we do that? Well, we take the comparison triangle $\overline{\Delta}$. Consider the comparison point $\overline{p_-}$ of $p_- = \gamma(t-\epsilon)$. Notice that $d_X(p_-, p_+) = 2\epsilon$. The CAT(0) inequality reads

$$2\epsilon = d_X(p_-, p_+) \leq d_{\mathbb{R}^2}(\overline{p_-}, \overline{p_+}).$$



But $\overline{p_-}$ and $\overline{p_+}$ lie on consecutive sides of a Euclidean triangle, and they are at distance ϵ from the common endpoint of such sides. The only way they can be at distance $\geq 2\epsilon$ is if the angle at the common vertex equals π. Hence, $\overline{\Delta}$ is degenerate, which implies that so is Δ. □

Here is one reason why the lemma is useful. A local isometric embedding is a map f between metric spaces X, Y so that every point in X has a neighborhood where f restricts to an isometric embedding.

Corollary 11.3.3. *A local isometric embedding between* CAT(0) *spaces is an isometric embedding.*

And in turn, we get:

Corollary 11.3.4. *Suppose that $\iota : X \to Y$ is a local isometric embedding of nonpositively curved spaces. Then the induced map at the level of fundamental groups is injective.*

Checking that a map between topological spaces induces an injection at the level of fundamental groups is tricky in general, so it's always good to have conditions for that to hold... We will see an application later.

Proof. The induced map $\tilde{f}$ between the universal covers of X and Y is also a local isometric embedding, whence an isometric embedding, and in particular it is injective. There are a few ways to conclude π_1-injectivity from here, using variable amounts of covering theory. for example one can argue as follows. Suppose that $\gamma : [0,1] \to X$ represents a non-trivial element of $\pi_1(X)$. Then any lift $\tilde{\gamma}$ to the universal cover of X has distinct endpoints, whence by injectivity $\tilde{f}(\tilde{\gamma})$ has distinct endpoints as well. However, $\tilde{f}(\tilde{\gamma})$ is a lift of $\iota(f(\gamma))$, which means that $\iota(f(\gamma))$ cannot be trivial in $\pi_1(Y)$ for otherwise all its lifts would be loops. □

Here is another very useful property of CAT(0) spaces.

Proposition 11.3.5. CAT(0) *spaces are contractible.*

We are not going to prove this property in detail. The idea is simple: One just uses geodesics to retract the space onto a point. It is possible to write down a well-defined map that does this because CAT(0) spaces are uniquely geodesic. Also, such map is going to be continuous because geodesics vary continuously, as they "vary at most as geodesics in $\mathbb{R}^2$".

The proposition turns out to be an actually useful tool to show that certain spaces are contractible. For example, it has been used to show the existence of manifolds with exotic properties, as in the theorem below. A manifold M is *aspherical* if its universal cover is contractible, or, equivalently (even though this is non-trivial!) if its higher homotopy groups $\pi_i(M)$, $i \geq 2$, are all trivial.

Theorem 11.3.6. *[Davis] For every $n \geq 4$, there exist closed aspherical n-manifolds whose universal cover is not homeomorphic to $\mathbb{R}^n$.*

The manifolds constructed by Davis do not admit metrics of nonpositive curvature, but still the proof involves CAT(0) spaces: The way that asphericity is shown is by showing that the universal cover is homotopy equivalent to some CAT(0) space (that is not a manifold).

11.4 Groups acting on CAT(0) spaces

We will say that a group is CAT(0) if it acts properly and coboundedly on a proper CAT(0) space.

Basic examples of CAT(0) groups include: $\mathbb{Z}^n$, free groups, fundamental groups of closed manifolds of nonpositive sectional curvature, products of CAT(0) groups. We will see another important example, right-angled Artin groups, in the chapter on CAT(0) cube complexes.

Just as an example, we state some of the many good properties that CAT(0) groups have:

Theorem 11.4.1. CAT(0) *groups are finitely presented, have solvable word problem, and have at most quadratic Dehn function.*

The idea for why the Dehn function is at most quadratic is that one can construct fillings of loops in CAT(0) spaces by "coning over a point" using geodesics as in Figure 11.2.

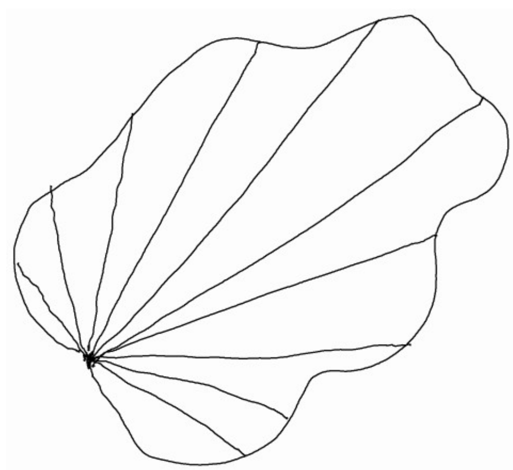


Figure 11.2

In whichever way you choose to make sense of this, the area of the disk spanned by such geodesics is at most (comparable with) the square of the length of the loop.

11.5 Connections with hyperbolicity

The main connection between CAT(0) and hyperbolic geometry is given by the following result.

Theorem 11.5.1. *A proper* CAT(0) *space* X *whose isometry group acts coboundedly is hyperbolic if and only if* X *does not contain isometrically embedded copies of* $\mathbb{R}^2$.

So, there's this very obvious obstruction to hyperbolicity, containing an isometrically embedded copy of $\mathbb{R}^2$, that in this case turns out to be the only obstruction. That's good.

If we have a CAT(0) group, then the CAT(0) space it acts on satisfies the requirements of the theorem. Keeping into account Milnor-Švarc Lemma, we get:

Corollary 11.5.2. *The* CAT(0) *group* G *is hyperbolic if and only if it does not contain a quasi-isometrically embedded copy of* $\mathbb{R}^2$.

It is an open question whether or not all hyperbolic groups are CAT(0) (!). Many, probably most people in the field believe that this is not true, though.

Finally, there is also a notion of CAT(-1) space. The definition is similar to that of a CAT(0) space, except that one considers comparison triangles in $\mathbb{H}^2$ instead of $\mathbb{R}^2$. Due to the hyperbolicity of $\mathbb{H}^2$, CAT(-1) spaces are hyperbolic.

11.6 Nonpositively curved complexes

In this section we consider the following

Question: When can one check if a given metric on a compact cell complex is nonpositively curved?

We like compact nonpositively curved cell complexes because their fundamental groups are CAT(0), as they act on the universal cover of the cell complex.

In general, it is very hard to decide, given a metric celle complex, if it is nonpositively curved or not. However, there are two classes of examples in which nonpositive curvature can be checked easily.

We analyse the first one now, while the second one is the topic of the next chapter.

Definition 11.6.1. A polygonal complex is a 2-dimensional metric cell complex where cells are (isometrically) identified with convex polygons in $\mathbb{R}^2$ and the cells are glued isometrically along sides.

Here are some (topologically planar) examples:

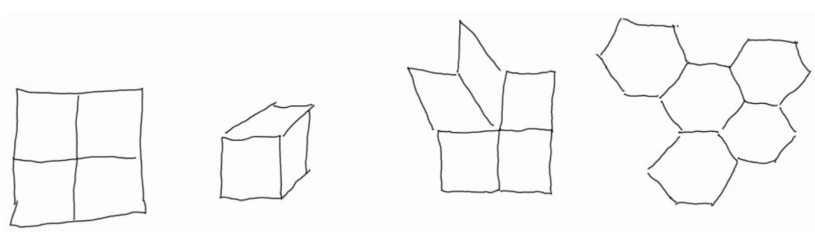


Figure 11.3

Notice that inside the polygons, the local CAT(0) condition is clear because any point has a neighborhood isometric to a convex subset of $\mathbb{R}^2$. It is also not very hard to see that there's no problem in the interior of the edges, as what you're doing there locally is just gluing convex parts of $\mathbb{R}^2$ along segments, so this is fine by Theorem 11.2.1. So, troubles can only happen at vertices.

We are about to define an object, the link, that captures the local geometry around a vertex, and we will be able to check nonpositive curvature by looking at links.

Let v be a vertex of a polygonal complex. Then the *link* Lk_v at v is the metric graph with

- a vertex corresponding to each half-edge emanating from v,
- edges corresponding to corner of polygons at v,
- the length of each edge equal to the angle (in radians) of the corresponding corner of polygon.

As usual, pictures are more illuminating:

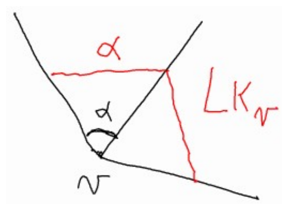


Figure 11.4

And here we go:

Theorem 11.6.2. *A compact polygonal complex is* CAT(0) *if and only if the length of any loop in any link is at least* 2π*.*

Here are loops in links from the examples drawn in Figure 11.3 and their lengths computed under the assumption that all drawn polygons are regular.

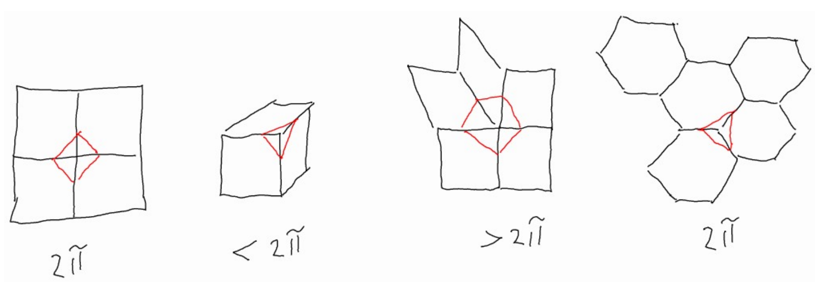
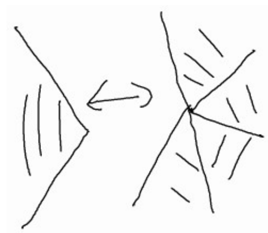


Figure 11.5

Notice that all the complexes above are nonpositively curved except for the second one. The slogan is "you are nonpositively curved if around each point there's at least as much stuff as in $\mathbb{R}^2$".

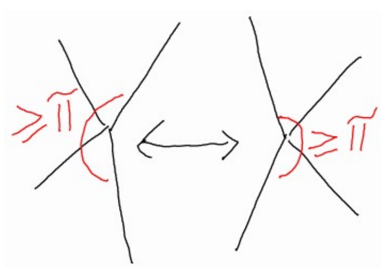
Let me try to justify why the condition is sensible, without giving a complete proof.

Let us consider all the (corners of) polygons at a given point that correspond to a loop in the link, and let us add them one by one. When we have only one polygon, we are fine. Now maybe we have to glue the next one to one side of the first one. We're still CAT(0), because of Theorem 11.2.1 (the one about gluing CAT(0) spaces along convex subspaces). We can keep going, except that we get into troubles when we glue in the last polygon.



You see, the last polygon gets glued along 2 sides, not just one, and the union of those two sides is not convex, as there are shortcuts in the inside.

What we want to do, and we can do it if and only if we have the condition on links stated in the theorem, is rearranging the gluing procedure (after subdivision of the polygons if necessary) in such a way that the final gluing looks like this:



Having angles of at least π on both sides guarantees that we are gluing along a convex subset, and then we are happy.

Chapter 12

CAT(0) cube complexes

A cube complex is just a cell complex obtained gluing Euclidean cubes isometrically along faces. As it turns out, it is easy to check when a cube complex is CAT(0) or nonpositively curved (the condition one needs to check has something to do with subcomplexes looking like corners of cube actually being corners of cubes).

The theory of CAT(0) cube complexes is surprisingly rich and goes well beyond what one can do with CAT(0) geometry only. In particular, as we will see, it has been used to prove some big conjectures in low dimensional topology.

12.1 Definitions, nonpositive curvature and flag links

Definition 12.1.1. An *n-cube* is just a copy of $[-1,1]^n$, endowed with the metric inherited from $\mathbb{R}^n$. A *face* of an n-cube is the subspace obtained setting one coordinate to either 1 or -1. A *cube complex* is a metric cell complex where all cells are (isometric to) n-cubes, and such that the gluing maps are isometries between faces and cubes[1].

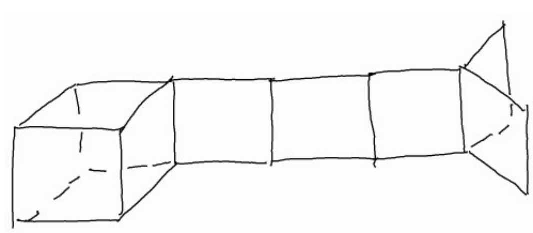


Figure 12.1: A cube complex

[1] Yeah, this doesn't sound right, but it's the formally correct formulation. It means that the $(n-1)$-cells are glued to faces of n-cells. In any case, hopefully the picture is clear enough...

Just as in the case of 2-complexes, the only obstructions to nonpositive curvature can arise at the vertices, everywhere else everything is fine. Once again, we define an object that encodes the local geometry around a vertex, which is called link in this case as well, and we will be able to check nonpositive curvature just by looking at the vertices.

In this case the link is a purely combinatorial object, there's no metric involved.

It is convenient (not necessary) to introduce the notion cubical subdvision of a cube complex, whereby we subdivide in the obvious way all n-cubes into 2^n smaller cubes, for each n.

Definition 12.1.2. Let v be a vertex of a cube complex X. The *link* Lk_v at v is the simplex-complex where

- the k-simplices are the $(k+1)-$cubes of the first cubical subdivision that contain v, and
- a simplex S_1 in Lk_v is a face of a simplex S_2 if the cube corresponding to the S_1 is a face of the cube corresponding to the second one.

Enough with this nonsense, here is what the link is:

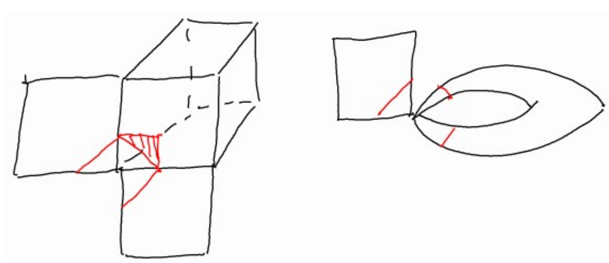


Figure 12.2

The picture on the right illustrates why one needs to be slightly careful and take the first barycentric subdivision: Cubes may be glued to themselves.

Definition 12.1.3. A *flag complex* is a simplicial complex so that any $n+1$ vertices span a simplex if and only if they are pairwise adjacent.

Recall that in a simplicial complex the simplices are uniquely determined by their vertices. For example, a loop or the complex with two vertices and two edges connecting them are not simplicial. Having said that, a simplicial complex is flag if there are no "empty simplices". Some examples are provided in Figure 12.3.

Figure 12.3

Theorem 12.1.4. *A cube complex is nonpositively curved if and only if all links are flag complexes. In particular, a cube complex is* CAT(0) *if and only if it is simply connected and all links are flag complexes.*

Here are some examples to illustrate the theorem.

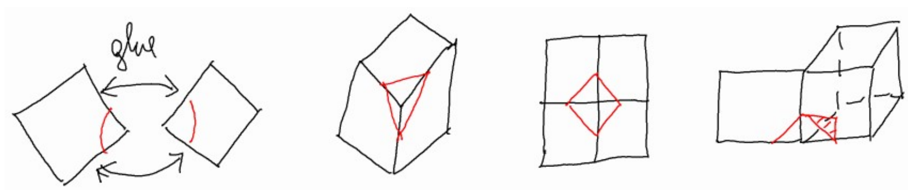


Figure 12.4

In the first picture, the link at v is not even simplicial, and in particular not flag. In the second example the link at v is an "empty simplex", so it is not flag, while the other examples are nonpositively curved.

The condition on links can also be rephrased as:

- If the k-cubes C, C' share k edges emanating from the same vertex, then $C = C'$ (this corresponds to links being simplicial).
- If X contains k 2-cubes all containing a vertex v and pairwise sharing an edge that emanates from v, then X contains a k-cube C containing all the given 2-cubes.

Number 2 is the condition that means: "If you see the corner of a cube, then the cube is there".

12.2 First examples

The simplest example of nonpositively curved cube complexes are graphs. In this case, the link is discrete and there's nothing to check.

Another simple example is the standard way of presenting the 2-torus as a quotient of a square.

In this case there is only one vertex and the link if a 4-cycle.

Another simple example is that of this surface:

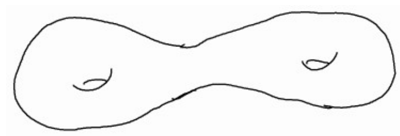


Figure 12.5

Such surface can be presented as a quotient of an octagon, and we can decompose such octagon into squares as suggested by the following picture:

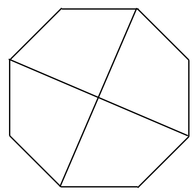


Figure 12.6

The corresponding cube complex is nonpositively curved. A similar construction works in higher genus as well.

12.3 Right-angled Artin groups...

In this section we discuss an important class of nonpositively curved cube complexes, and their fundamental groups.

Definition 12.3.1. Let Γ be a finite graph. The *right-angled Artin group* (RAAG) A_Γ with presentation graph Γ is the group given by the presentation

$$A_\Gamma = \langle \Gamma^0 | vw = wv \iff (v,w) \in \Gamma^1 \rangle,$$

where Γ^0 is the set of vertices of Γ and Γ^1 its set of edges.

In words, A_Γ is generated by the vertices of Γ and two vertices commute if they are connected by an edge. Simple, right?

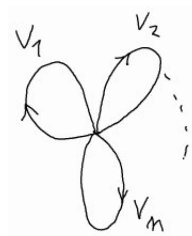
Basic examples include:

- $A_\Gamma = \mathbb{Z}^n$ if Γ is the complete graph on n vertices.
- $A_\Gamma = F_n$ if Γ has n vertices and no edges.
- $A_\Gamma = A_{\Gamma_1} * A_{\Gamma_2}$ if $\Gamma = \Gamma_1 \sqcup \Gamma_2$.
- *Exercise:* A_Γ is the direct product of A_{Γ_1} and A_{Γ_2} if Γ is... what?

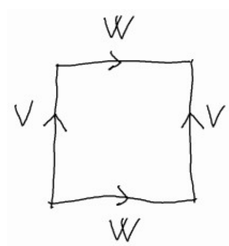
12.3.1 Salvetti complex

Let us now describe the *Salvetti complex* S_Γ associated to Γ, which is a nonpositively curved cube complex whose fundamental group is A_Γ.

Start with a bouquet of circles with one loop corresponding to each vertex of Γ:

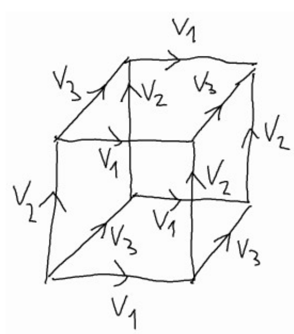


We also fix an orientation of edge loop. For the moment the fundamental group is just a free group on the vertices of Γ. We now proceed inductively. Whenever v, w are connected by an edge we glue in a 2-cube with the following gluing instructions:



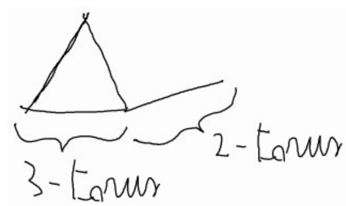
Notice that opposite edges get glued together and hence the image of the square in the Salvetti complex is a 2-torus.

Now we have the right fundamental group, but we might not be nonpositively curved yet (there might be "holes" that look like an empty cube). So, whenever we have vertices $v_0, \ldots, v_k$ that commute we glue in a k-cube with edges in the i-th direction marked by v_i, here is the picture for $k = 3$:



Of course, we glue the faces of such k-cube to the corresponding $(k-1)$-cubes that had been added at the previous step of the induction. In particular, opposite faces get identified and in particular the image of the k-cube in the Salvetti complex is a k-torus.

As a concrete example, if Γ looks like this:



Then the Salvetti complex is obtained gluing a 3-torus and a 2-torus along a loop.

12.4 ... and their subgroups

Right-angled Artin groups have several nice properties, here are some of them.

Theorem 12.4.1. *Let Γ be a finite graph. Then*

1. *A_Γ embeds in $GL(n, \mathbb{Z})$ for some n.*
2. *A_Γ is residually finite, i.e. for every non-trivial $g \in A_\Gamma$ there exists a finite group F and a homomorphism $\phi : A_\Gamma \to F$ so that $\phi(g) \neq 1$.*
3. *if Γ is not complete then A_Γ surjects onto $\mathbb{F}_2$.*

Notice that the first two properties are inherited by subgroups. The reason why we care about subgroups of RAAGs is the following.

Theorem 12.4.2. *[Haglund-Wise, 2009] Let X be a compact* special *cube complex. Then $\pi_1(X)$ embeds in some RAAG.*

We haven't defined special cube complexes just yet, but whatever that means they are not so special after all because of the following important result:

Theorem 12.4.3. *[Agol, 2012] If X is a compact nonpositively curved cube complex then $\pi_1(X)$ contains a finite index subgroup H that is the fundamental group of a special cube complex.*

The theorem is important because (a slight more general version of) it was the final step in the proof of the Virtual Fibering Conjecture, now Theorem, which says:

Theorem 12.4.4. *Let M be a closed hyperbolic* 3*-manifold. Then M has a finite sheeted cover N that fibres over the circle.*

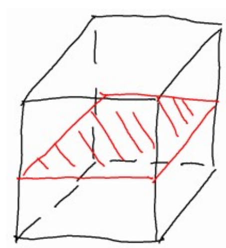
So, N has the following simple description: Take $S \times [0, 1]$, where S is some closed surface, and glue the boundary components using some homeomorphism. Hence, up to passing to a finite sheeted cover, any closed hyperbolic 3-manifold admits such a description.

The proof of the Virtual Fibering Theorem uses a criterion (due to Agol) that says, in simplified form, that M virtually fibres over the circle if its fundamental group embeds in some RAAG. Another key step is showing that fundamental groups of 3-manifolds can be "cubulated", so that one can use the two theorems quoted above.

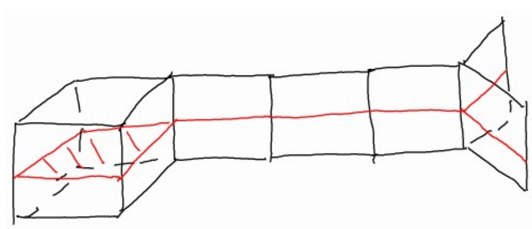
12.5 Special cube complexes

Time to define special cube complexes. The definition has something to do with the notion of hyperplane, that we are about to define.

A *midcube* is the subspace of a cube obtained setting one coordinate to 0.



A hyperplane is, informally speaking, what you get if you start from some midcubes and keep extending across faces to include other midcubes:



Formally, one defines two edges in a cube complex to be *directly parallel* if they intersect the same midcube, while two edges e, e' are *parallel* if there is a chain $e = e_0, \ldots, e_n = e'$ of edges so that e_i, e_{i+1} are directly parallel. Finally, a hyperplane is a parallelism class of edges. Not that we really care, of course.

The definition of special cube complex is fine-tuned to make the proof of Theorem 12.4.2 work. Certain "pathologies" of hyperplanes would prevent the proof from working, and they are exactly the ones that are excluded by the definition of specialness. We now give the definition, but you might want to read the proof of Theorem 12.4.2 below instead and come back here when a property of hyperplanes is needed in the proof.

Definition 12.5.1. A nonpositively curved cube complex is special if the hyperplanes if the following configurations do NOT arise:

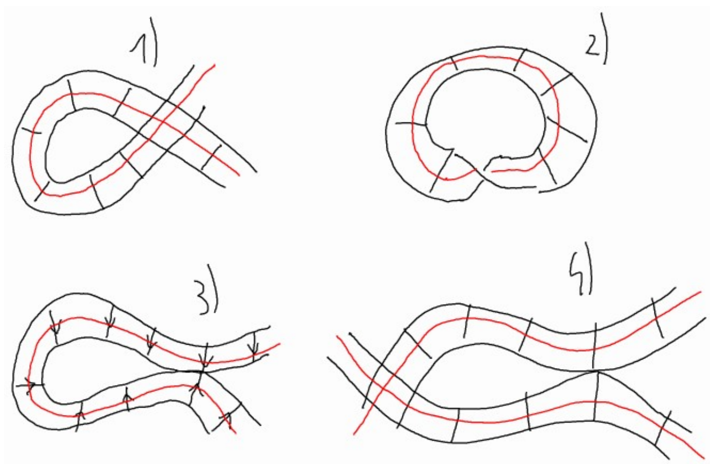


Figure 12.7: 1) hyperplanes are embedded, 2) hyperplanes are 2-sided, 3) hyperplanes do not self-osculate, 4) hyperplanes do not inter-osculate

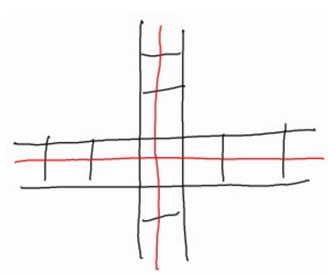
In order to do things formally, one would have to translate the pictures into the language of parallel edges. Let us do this just for the first example: There does not exist a cube so that edges intersecting distinct midcubes are parallel.

Let us know recall Theorem 12.4.2 and sketch-prove it.

Theorem 12.5.2. *Let X be a compact* special *cube complex. Then $\pi_1(X)$ embeds in some RAAG.*

Outline of proof. What we are really going to prove is that there exists a local isometric embedding of X into some Salvetti complex. Then we can invoke Corollary 11.3.4.

The graph Γ that we are going to use is the *crossing graph* of X. The vertices of such graph are the hyperplanes, and there is an edge connecting two hyperplanes if and only if they cross, meaning that they intersect, meaning this:



As hyperplanes of X are two-sided (property 2), we can orient coherently all edges dual (formally: contained) in a given hyperplane. Just pick any orientation of one edge, put the obivous orientation on all the edges dual to the same midcube and keep extending.

We are now ready to define a map from the 1-skeleton of X to the 1-skeleton of S_Γ, which is just a bouquet of oriented loops each corresponding to a hyperplane of X.

The map is defined as follows. The image of an edge e of X is the loop in S^1_Γ corresponding to the (only) hyperplane dual to e, and e is mapped isometrically and orientation-preservingly onto such edge.

If this map is going to be a local isometric embedding it'd better be locally injective. But what would it mean for two edges in X emanating from the same vertex to be mapped onto the same edge in S_Γ? It would mean that two such edges are contained in the same hyperplane. But this cannot happen for edges contained in a common square because hyperplanes are embedded (property 1). Also, it cannot happen for edges not contained in a common square because of no interosculation (property 3). So, it can never happen. (There is a subtlety about orientation in the second case, can you see it?)

Now, there is a combinatorial characterisation of local isometries. If you are willing to believe me, what you need to check (once you have local injectivity) is exactly that the following obvious obstruction to being a local isometric embedding does not arise at any vertex v: Two edges emanating from v that are not contained in a common square get mapped to edges contained in a common square. And this is exactly what property 4 prevents.

In order to conclude the proof, one needs to check that the map we defined on the 1-skeleton extends to higher skeleta as well, and we will only mention that this uses the flag condition. □

Chapter 13

Cubulation

In this chapter we describe the most important way to construct CAT(0) cube complexes. The main idea is that the structure of a CAT(0) cube complex can is entirely encoded by the combinatorics of its hyperplanes, and in turn the way in which hyperplanes cut up a CAT(0) cube complex can be axiomatised in terms of the notion of wall.

In this chapter it will be most convenient to regard a hyperplane as an actual subspace of the CAT(0) cube complex rather than as a parallelism class of edges, and hence we will do so.

13.1 Properties of hyperplanes

Let us start with stating the main properties of hyperplanes.

Theorem 13.1.1. *Let H be a hyperplane in the* CAT(0) *cube complex X. Then*

1. *H is embedded,*
2. *H is convex,*
3. *$X \backslash H$ has exactly two connected components.*

Let me also remark that H is in a natural way a CAT(0) cube complex itself. The dimension of H (meaning the maximal dimension of cubes in H) is strictly smaller than the dimension of X. This is useful because sometimes it can be used to prove statements about CAT(0) cube complexes using induction on dimension.

We are actually not going to use the theorem, but I thought it was worth stating it anyway to motivate what we are going to do.

Just to give an idea about how one can prove (part of) it, let me recall that local isometric embeddings between CAT(0) spaces are (global) isometric embeddings (Lemma 11.3.3). Hence, you get 1) and 2) if you show that the inclusion of H into X is a local isometric embedding, and this shouldn't sound too unreasonable as after all H is built up from midcubes, that are convex subspaces of cubes.

In order to convince yourself of the third property, which is actually the focus of this chapter, you should first think of the *carrier* $N(H)$ of the hyperplane, which is just the union of the cubes that the hyperplane goes through. It shouldn't be hard to believe that $N(H)$ can be naturally identified with $H \times [-1, 1]$, just like a cube can be identified with one of its midcubes times $[-1, 1]$. In particular, H locally disconnects X into two connected components. One possible way to conclude at this point is given in the following exercise.

Exercise. Let C be a proper convex subset of a CAT(0) space X.

1. Show that for each $x \in X$ there exists a unique point $\pi_C(x)$ in C that minimises the distance from x, $d(x, \pi_C(x)) = d(x, C)$. (Properness helps but it's not strictly needed.)

2. Show that $\pi_C : X \to C$ is 1-Lipschitz, and in particular continuous.

3. Going back to our CAT(0) cube-complex X with a hyperplane H, convince yourself that $N(H)$ is convex and that for each $x \in X \backslash N(H)$, we have that $\pi_{N(H)}(x)$ lies in either $H \times \{1\}$ or $H \times \{-1\}$.

4. Use the previous point to show that $X \backslash H$ has two connected components.

13.2 From walls to cube complexes

Our aim is now to describe some structure one can put on a set that mimics the way hyperplanes chop up a CAT(0) cube complex.

Definition 13.2.1. Let X be a set. A *wall partition* (W^-, W^+) of X is just a pair of subsets so that $X = W^- \cup W^+$. In such case, we call $W = W^+ \cap W^-$ the *wall* and $\mathring{W}^\pm = W^\pm - W$ the *half-spaces*. We say that $p, q \in X$ are *separated* by the wall W if $p \in \mathring{W}^+$ and $q \in \mathring{W}^-$ or vice-versa.

Nothing too fancy, right? As you can probably guess, a space with walls is a set together with a collection of wall partitions (satisfying two mild conditions).

Definition 13.2.2. A *space with walls* is a set X together with a collection of wall partitions $\mathcal{W}$ so that

1. every point in X is contained only in finitely many walls,

2. every pair of point in X is separated only by finitely may walls.

Our motivating example of space with walls is of course a CAT(0) cube complex X together with the wall structure given by the partitions of X induced by its hyperplanes.

So, the first question is now: How does one recover the vertex set of the cube complex X from its hyperplanes?

The answer is that, given a vertex v of X, one can keep track of which side of each hyperplane v lies in. You can think of this as adding an arrow to each hyperplane pointing towards v:

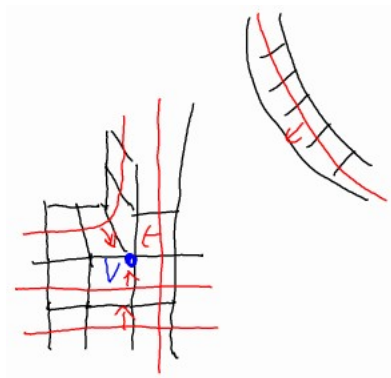


Figure 13.1

This motivates the following definition:

Definition 13.2.3. An *orientation* $o(W)$ on a wall W is just a choice of one of W^+, W^-, i.e. either $o(W) = W^+$ or $o(W) = W^-$. Given an orientation $o(W)$ on the wall W for the set X, we say that W is *oriented towards* $x \in X$ if $x \in o(W)$. A *coherent orientation* on a space with walls $(X, \mathcal{W})$ is an orientation on each wall so that:

1. $o(W) \cap o(W') \neq \emptyset$ for each $W, W' \in \mathcal{W}$.
2. for any $x \in X$ there are only finitely many walls W *not* oriented towards x.

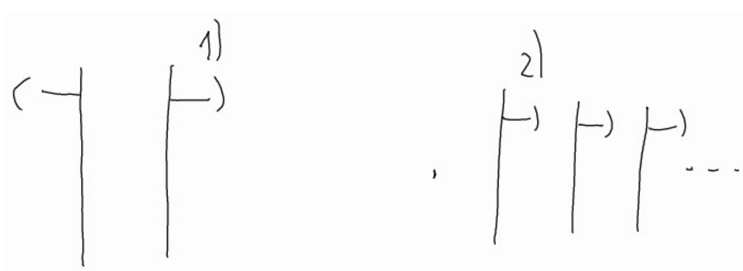


Figure 13.2: NON-allowed configurations

It is reasonably clear that if we orient all hyperplanes of a *proper* $CAT(0)$ cube complex X towards one of its points as we did above, then the two conditions hold, so we can think of a vertex in X as an orientation on the hyperplanes instead.

Now, the last observation: If x, y are vertices of the $CAT(0)$ cube complex X and they are connected by an edge e, then the corresponding orientations differ exactly on one hyperplane, namely the one dual to e.

We now have everything we need to construct a $CAT(0)$ cube complex from a space with walls.

Definition 13.2.4. Let $(X, \mathcal{W})$ be a space with walls. Then its *dual cube complex* $C = C(X, \mathcal{W})$ is the cube complex with

- vertices $\leftrightarrow$ coherent orientations on $\mathcal{W}$,
- edges $\leftrightarrow$ orientations that differ on exactly one wall, and
- n-cubes $\leftrightarrow$ 1-skeleta of n-cubes.

The last item means that we plug in an n-cube wherever we see a subcomplex of the 1-skeleton of C isomorphic to the 1-skeleteon of an n-cube.

And the theorem is...

Theorem 13.2.5 (Sageev)**.** *If $(X, \mathcal{W})$ is a space with walls, then its dual cube complex is $CAT(0)$.*

It is important to remark that the construction of the dual cube complex is natural in the sense that if $(X, \mathcal{W})$ is a space with walls and you have a group G acting on X preserving $\mathcal{W}$, then G is going to act on the dual cube complex as well. Hence, the theorem gives a way to construct actions on $CAT(0)$ cube complexes.

13.3 Examples

We already discussed the example of cube complexes during the discussion leading to the definition of the dual cube complex, but let us analyse a concrete example.

1. Consider the standard cubulation of $\mathbb{R}^2$. The walls/hyperplanes naturally come in two families, corresponding to horizontal and vertical lines. A coherent orientation is uniquely determined by a horizontal and a vertical "switch point", meaning the following. Suppose that you want a coherent orientation just on the vertical lines, for the moment. You cannot have the configuration in Figure 13.2-(1).

   Also, not all hyperplanes can be oriented in the same direction, for otherwise infinitely many of them will point away from the origin.

   This forces the orientation to look like in Figure 13.3.

   The switch point is where hyperplanes switch from being oriented to the right to being oriented to the left.

   A similar argument applies to the horizontal lines, and you can check that you can choose orientations on horizontal and vertical lines independently.

   Here are a few more examples.

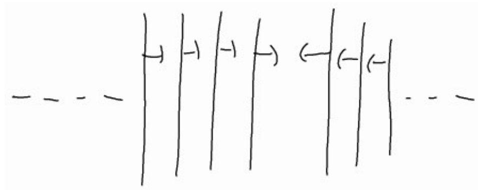
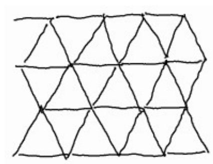
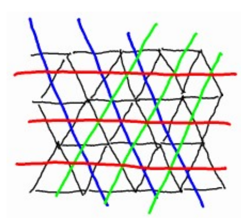


Figure 13.3

2. Let G be the group of symmetries of the tessellation of $\mathbb{R}^2$ by equilateral triangles:



$\mathbb{R}^2$ has a G-invariant wall structure $\mathcal{W}$ with three families of walls as suggested by the following picture:



In each family of wall an orientation is determined by a "switch point", and the switch points can be chosen independently, like in the previous case. Hence, in this case the dual cube complex C is $\mathbb{R}^3$ with the standard cubulation.

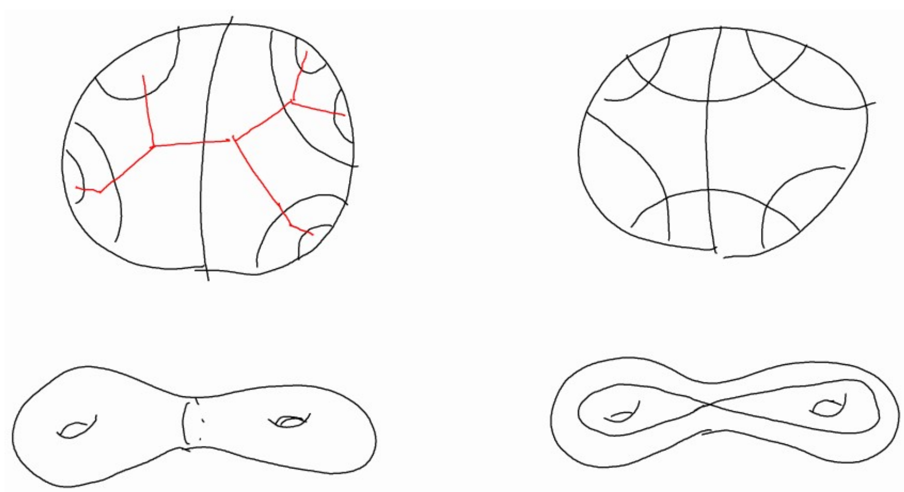
The action of G on C is proper but cannot possibly be cobounded (G is "2-dimensional" for any meaningful definition of dimension).

Actually, there are conditions for the action on the dual cube complex to be proper and/or cobounded, but we will not go into this (even though it's very important).

3. Walls in small cancellation groups. Let $\langle S|R\rangle$ be a $C'(1/6)$ presentation, say with all relators of even length for simplicity. Here is an idea to construct walls in the Cayley complex. Define a mid-cell to be a line that goes from the midpoint of one edge in a relator to the midpoint of the opposite edge.

As it turns out, if you start from a midcell and maximally extend through the edges, you get an embedded graph that separates your Cayley complex in two connected components. So, you can define a wall structure in this way. If the presentation is finite, then the dual cube complex is finite dimensional (this is related to the fact that there's a bound on the number of walls that pairwise intersect). Also, once again if the presentation is finite, $\langle S|R\rangle$ acts properly and coboundedly on the dual cube complex.

4. If you have a closed (not null-homotopic) curve on a surface, say a closed surface of genus 2, then in the universal cover, $\mathbb{H}^2$, one sees a family of lines each of which separates $\mathbb{H}^2$ into two connected components. The more complicated the self-intersections of the curve are, the more complicated this pattern of lines is.



   The dual cube complex in the first case is a tree, while in the second case it has dimension 2. If you have more complicated curves, then you can have higher dimensional complexes as well...

5. Let me briefly mention that the analogue in dimension 3 of the previous example is considering a π_1-injective surface, or a family of such surfaces, in a 3-manifold. This is the construction that eventually leads to the Virtual Fibering Theorem 12.4.4. One needs “sufficiently many” surfaces to get a proper cobounded action on a CAT(0) cube complex, this is due to Bergeron-Wise. The existence of “sufficiently many” surfaces in a hyperbolic 3-manifold is instead due to Kahn-Markovic.