A Short course in geometric group theory

Notes for the ANU Workshop January/February 1996

Walter D. Neumann and Michael Shapiro

§0 Introduction: Dehn's 3 Questions

1) The word problem.
2) The conjugacy problem.
3) The isomorphism problem.
4) Boone and Novikov's answer: No!
5) The geometric viewpoint: Cayley graphs.
6) A group G has soluble word problem if and only if its Cayley graph can be constructed recursively.
7) Dehn's solution to the word problem for hyperbolic surface groups.

§1 Cayley graphs and hyperbolicity

1) Given a group G and a finite generating set $\mathcal{G}$, the following are equivalent:
   i) Triangles in Γ are thin.
   ii) Triangles in Γ are thin in a parameterized way.
   iii) Bigons in Γ are thin.
   iv) G has a linear isoperimetric inequality.
   v) G has a sub-quadratic isoperimetric inequality .
   vi) G has a Dehn's algorithm.
   vii) Geodesics in Γ exhibit exponential divergence.
   viii) Geodesics in Γ exhibit super-linear divergence.
2) Milnor's theorem: If G acts co-compactly and discretely by isometries on a geodesic metric space X, then the Cayley graph of G is quasi-isometric to X.
3) If G acts co-compactly and discretely by isometries on a negatively curved geodesic metric space X, then G is hyperbolic.

§2 Elementary properties of hyperbolic groups

1) Behaviour of geodesics and quasi-geodesics: progression in geodesic corridors.
2) Hyperbolicity is a quasi-isometry invariant.
3) Finite presentation, solubility of the word and conjugacy problems.
4) Finitely many conjugacy classes of torsion elements.
5) The Rips complex and FP_∞.
6) The boundary.

§3 Isoperimetric inequalities

1) Area of a word via products of conjugates of relators.
2) Area of a word via Van Kampen diagrams.
3) Dehn's functions and isoperimetric functions.
4) A group has soluble word problem if and only if it has a recursive Dehn's function if and only if it has a sub-recursive Dehn's function.
5) Equivalence relation and ordering for isoperimetric functions.
6) Consequently solubility of the word problem is a geometric property.

§4 The JSJ decomposition

1) Splittings of hyperbolic groups.

§5 Regular languages, automatic, bi-automatic and asynchronously automatic groups

1) Definitions.
2) The fellow traveler property.
3) The seashell: finite presentation, quadratic isoperimetric inequality and quadratic time word problem.
4) Closure properties.
   i) Free products.
   ii) Direct products.
   iii) Free factors.
   iv) Finite index subgroups.
   v) Finite index supergroups.
   vi) HNN extensions and free products with amalgamation over finite subgroups.
   vii) Bi-automatic groups are closed under central quotients and direct factors.
5) Famous classes of automatic groups.

i) Hyperbolic groups including: free groups, finite groups, most small cancellation groups, fundamental groups of closed negatively curved manifolds and other negatively curved spaces.
ii) Small cancellation groups.
iii) Fundamental groups of geometric three manifolds except for those containing a nil or solv manifold as a connected sum component.
iv) Coxeter groups.
v) Mapping class groups.
vi) Braid groups and more generally, Artin groups of finite type.
vii) Central extensions of hyperbolic groups.
viii) Many amalgams of hyperbolic groups along rational subgroups.
ix) Many groups that act on affine buildings.

6) Regular languages, cone types and the falsification by fellow traveler property.
7) Bi-automatic groups
   i) Hyperbolic groups are geodesically bi-automatic.
   ii) Bi-automatic groups have soluble conjugacy problem.
8) Asynchronous automaticity

§6 Equivalence classes of automatic structures

1) The asynchronous fellow traveller property, rationality and bi-automaticity.

§7 Subgroups of automatic groups

1) Let L be a (synchronous or asynchronous) automatic structure for G, and suppose $H < G$. Then the following are equivalent:
   i) H is L-rational.
   ii) H is L-quasi-convex.
   iii) In the case where G is hyperbolic and L is synchronous, these are equivalent to: H is quasi-geodesic in G.
2) A rational subgroup of an automatic (bi-automatic, asynchronously automatic, hyperbolic) group is automatic (bi-automatic, asynchronously automatic, hyperbolic).
3) Centers and centralizers of bi-automatic groups are rational, hence bi-automatic.
   i) Consequently a hyperbolic group does not contain a $\mathbb{Z}^2$ subgroup.

§8 Almost convexity

1) Definition.
2) Building the Cayley graph.
3) Almost convexity is not a group property.

§9 Growth functions and growth rates

1) Finite cone types implies rational growth function.
2) Gromov's theorem: polynomial growth implies virtually nilpotent.

0. Introduction

Geometric group theory is the study of groups from a geometric viewpoint. Much of the essence of modern geometric group theory can be motivated by a revisitation of Dehn's three decision-theoretic questions, which we discuss below, in light of a modern viewpoint. This viewpoint is that groups may be profitably studied as geometric objects in their own right. This connection between algorithmic questions and geometry is at first surprising, and is part of what makes the subject attractive.

As Milnor's theorem (see below) teaches us, the geometry exists both in the group itself and in the spaces it acts on. This is another powerful motivation for the subject, and was historically one of the turning points.

0.1–4 Dehn's problems.

Early in this century Dehn proposed three seminal questions **[D]**.

Suppose we have a group G given in some way, and a set $\mathcal{G}$ of generators for the group. The ***word problem*** asks if there is a procedure to determine whether two words w_1 and w_2 in these generators represent the same element of G. Since we may look at $w_1^{-1}w_2$, it is equivalent to ask for a procedure to decide if a word w represents the identity.

Notice that if $\mathcal{G}$ and $\mathcal{G}'$ are two finite generating sets for G then the word problem is soluble with respect to $\mathcal{G}$ if and only if it is soluble with respect to $\mathcal{G}'$. For suppose we are given a word $w = a_1 \dots a_k$ in the letters of $\mathcal{G}'$. Each letter a_i has the same value in G as some word $w_i = b_{i1} \dots b_{ij(i)}$ in the letters of $\mathcal{G}$. Thus it requires only a finite look-up table to translate a word in the letters of $\mathcal{G}'$ into one in $\mathcal{G}$.

However, there is less here than meets the eye. We have said nothing about how to find the finite look-up table $a_i \mapsto b_{i1} \dots b_{ij(i)}$. Thus while we have demonstrated the *existence* of an algorithm for the word problem in $\mathcal{G}'$ from the existence of one in $\mathcal{G}$, we may have no way of *finding* that solution.

The ***conjugacy problem*** asks if there is a procedure to determine whether the elements g_1 and g_2 of G represented by two words w_1 and w_2 are conjugate in G, that is, does there exist $g \in G$ with $g_2 = g^{-1}g_1g$. Note that if such a procedure exists, then, taking $g_2 = 1$, we solve the word problem also.

The ***isomorphism problem*** asks for a procedure to determine whether two groups are isomorphic. The groups are usually assumed to be given by presentations (a "presentation" is a collection of generators for the group plus some sufficient collection of relations among these generators).

Dehn's three questions are remarkable in that they precede by many years the formalization of "procedure" as "algorithm" by Church, Turing, et al. and by two more decades the resolution of the word problem by Boone and Novikov in the 1950's. They show that there is a finitely presented group with recursively unsolvable word problem. The unsolvability of the conjugacy problem is immediate, and that of the isomorphism problem also follows. For more details see C.F. Miller's notes for this Workshop.

0.5 Cayley graphs

There is a standard method for turning a group and a set of generators into a geometric object. Given the group G and generating set $\mathcal{G}$ we produce the directed labeled graph $\Gamma = \Gamma_{\mathcal{G}}(G)$. The vertices of Γ are the elements of G and we draw a directed edge from g to ga with label a for each $g \in G$ and $a \in \mathcal{G}$. We will refer to this edge as (g, a, ga). This graph is called the ***Cayley graph***. It comes equipped with a natural left action of G. G acts on vertices by left-multiplication and $h \in G$ carries the edge (g, a, ga) to (hg, a, hga).

By forgetting directions on edges and making each edge have length 1, we can turn Γ into a metric space. This induces the ***word metric*** on the vertex set G:

$$d(g, g') = \min\{\mathrm{len}(w) : g' = gw \text{ in } G\}.$$

This metric depends of course on the chosen set of generators $\mathcal{G}$. But we will see below that up to "quasi-isometry" it does not depend on this choice (if the generating set is finite) and that many geometric properties depend only on G.

We use $\mathcal{G}^*$ to denote the set of all words on letters of $\mathcal{G}$, including the empty word ϵ and for $w \in \mathcal{G}^*$ we use $\overline{w}$ to denote its value in G. Any $w \in \mathcal{G}^*$ represents a unique edge path in Γ based at the identity 1. Using the same symbol w for the word and the path will not lead to confusion. We parameterize w by arc-length, so it is a map from $[0, \mathrm{len}(w)]$ to Γ. It is often convenient to extend this map to $[0, \infty)$ by defining $w(t) = \overline{w}$ for $t > \mathrm{len}(w)$. The path w is a ***geodesic*** if $d(1, \overline{w}) = \mathrm{len}(w)$.

0.6 The word problem and Cayley graphs

The word problem has a simple characterization in terms of the Cayley graph. A word $w \in \mathcal{G}^*$ labels a path starting at the identity and ending at the value $\overline{w} \in G$. Evidently, a word represents the identity if and only if it is a closed loop. It follows that

Theorem. $G = \langle \mathcal{G} \rangle$ *has soluble word problem if and only if there is an algorithm capable of constructing any finite portion of* $\Gamma_{\mathcal{G}}(G)$.

For suppose we are in possession of such an algorithm, and we are given the word w. The path w lies entirely inside the ball of radius $\mathrm{len}(w)$ around the identity. We use our algorithm to construct this ball and then follow w to see if it returns to the identity.

On the other hand, suppose we are given an algorithm to solve the word problem in G. To construct the ball of radius n around the identity, we enumerate the words of length less than or equal to n in $\mathcal{G}^*$. We then use our algorithm for the word problem to determine which of these are equal in G. Equality in G is an equivalence relation on $\mathcal{G}^*$. Pick a representative (say a shortest representative) in each equivalence class. Now for each $a \in \mathcal{G}$ and each pair g, g' of representatives, use word problem algorithm to determine if (g, a, g') is an edge.

0.7 Dehn's solution to the word problem for hyperbolic surface groups.

The previous algorithms are wildly impractical in most situations. There is a beautiful and highly efficient solution to the word problem for hyperbolic surface groups. Let us start by describing an inefficient solution to the word problem in $\mathbb{Z}^2$.

If we take $\mathbb{Z}^2$ in the standard presentation $\langle x, y \mid xyx^{-1}y^{-1} \rangle$, the Cayley graph embeds in the Euclidean plane as the edges of the tessellation by squares. We can see this embedding as an expression of the fact that $\mathbb{Z}^2$ acts by isometries on the Euclidean plane. The quotient of the Euclidean plane by this action is the torus. Now we are entitled to see the tessellation of the plane by squares as being the decomposition of the plane into copies of a fundamental domain for this action. However, there is a little piece of sleight-of-hand going on here. To see this, let's start with the torus.

If we cut the torus open along two curves, it becomes a disk, in fact, if we choose the two curves correctly, it is a square. That is, the torus is the square identified along its edges. If we look for the generators of the fundamental group, they are dual to the curves we have cut along. That is, if we take a base point in the middle of the square, x is the curve which starts at the base point, heads towards the right edge of the square, reappears at the left edge and continues back to the base point. Likewise, y is the vertical path which rises to the top edge and reappears at the bottom. Thus the fundamental group is generated by the act of crossing either of the cut curves.

Thus, after choosing a base point, the natural relation between the tessellation of the plane and the embedding of the Cayley graph into the plane is that they are ***dual***. That is, there is a vertex of the Cayley graph in the center of each copy of the fundamental domain, and there is an edge of the Cayley graph crossing each edge of the tessellation.

The reason it was not immediately obvious that the tessellation and the Cayley graph are two different things is that the tessellation of the plane by squares is ***self-dual***. That is, if we start with the tessellation of the planes by squares and replace each vertex by a 2-cell and each 2-cell by a vertex (thus getting new edges crossing our old edges) the result is another tessellation of the plane by squares. Contrast this with the tessellation of the plane by equilateral triangles which is dual to the tessellation of the plane by regular hexagons.

Now as we have seen, finding the Cayley graph for $\mathbb{Z}^2$ solves the word problem for $\mathbb{Z}^2$, but it does not lead to a particularly efficient solution.

This situation changes if we turn to the fundamental group of a hyperbolic surface, i.e., a surface of genus 2 or more. If we wish to cut open a surface of genus 2, we will need 4 curves, and when we have cut it open, we will have an octagon as our fundamental domain. Likewise, looking at the vertex where our cut curves meet shows that we will want to tessellate *something* with 8 octagons around each vertex. This suggests that we would like regular octagons with interior angles of 45°. We can have this if we choose to work in the hyperbolic plane. In fact, the hyperbolic plane can be tessellated by regular octagons with interior angles of 45°.

Once again, the Cayley graph is dual to the tessellation, and this tessellation is also self-dual, so the Cayley graph embeds as the edges of this tessellation. In fact, reading the labels on an octagon tells us that our fundamental group has the presentation

$$G = \langle x_1, y_1, x_2, y_2 \mid [x_1, y_1][x_2, y_2] \rangle.$$

Given a sub-complex X of this tessellation, we define $Star(X)$ to be X together with any fundamental domains which meet X. Now suppose we start with a vertex $*$ of this tessellation (identified with the identity) and consider $Star^n(*)$ for successive values of n. We can then observe that each fundamental domain that meets the boundary of $Star^n(*)$ meets it in at least 5 of its 8 edges.

Suppose now that we are given a word w which represents the identity. We consider w as a path based at $*$. There is a smallest n so that w lies entirely in $Star^n(*)$. If $n = 0$ then w is the empty word. If $w \neq 0$, then there is some portion u of w which lies in $Star^n(*)$, but not in $Star^{n-1}(*)$. Suppose u is a maximal such portion. Then one of 2 things happens. Either:

1) u returns to $Star^{n-1}(*)$ along the same vertex by which it left it, in which case w is not reduced or
2) u travels along the ≥ 5 edges of some 2-cell F which meet the boundary of $Star^n(*)$. Call this portion u_0.

In the first case, w can be reduced in length by deleting a subword of the form aa^{-1}, producing a new word w_1 with the same value in G.

In the second case, we take u_1 to be the path so that $u_0 u_1$ forms the boundary of F. We now have $u_0 u_1$ evaluating to the identity, so u_1^{-1} and u_0 have the same value in G. Further, $\mathrm{len}(u_1) \leq 3 < 5 \leq \mathrm{len}(u_0)$.

If we work with a surface of genus greater than 2, the numbers change, but the argument stays the same. Thus we have proven:

Theorem (Dehn). *Let G be the fundamental group of a closed hyperbolic surface. Then there is a finite set of words $\mathcal{D} = \{v_1, \ldots, v_k\}$ with each v_i evaluating to the identity, so that if w is a word representing the identity, then there is some $v_i = u_0 u_1 \in \mathcal{D}$ so that $w = \alpha u_0 \beta$ and $\mathrm{len}(u_1) < \mathrm{len}(u_0)$.*

This gives a very efficient solution to the word problem. Given a word w we look for an opportunity to shorten it and do so if we can. After at most $\mathrm{len}(w)$ such moves we either arrive at the empty word or we have no further opportunities to shorten our word. If the first happens, we have shown that w represents the identity. If the second happens, we have shown that it does not.

1.1 Hyperbolicity

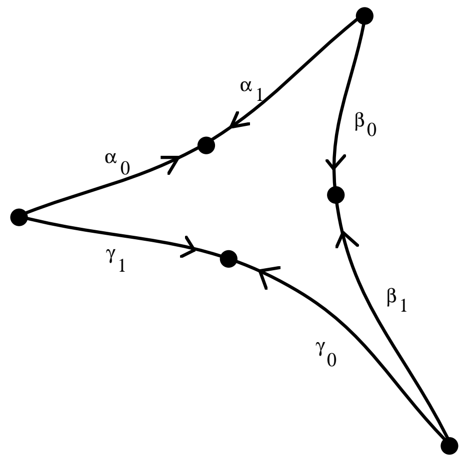
Let G be a finitely generated group with finite generating set $\mathcal{G}$ and let $\Gamma = \Gamma_{\mathcal{G}}(G)$ be the corresponding Cayley graph.

i). We say that Γ has ***thin triangles*** if there exists a δ such that if α, β, and γ are sides of a geodesic triangle in Γ then α lies in a δ-neighbourhood of $\beta \cup \gamma$.

ii). We can give a parameterized version of the above condition. We first note the following. If $\Delta = \alpha\beta\gamma$ is a geodesic triangle, then the sides of Δ decompose (as paths parameterized by arclength) as $\alpha = \alpha_0\alpha_1^{-1}$, $\beta = \beta_0\beta_1^{-1}$, $\gamma = \gamma_0\gamma_1^{-1}$ so that

$\mathrm{len}(\alpha_1) = \mathrm{len}(\beta_0)$, $\mathrm{len}(\beta_1) = \mathrm{len}(\gamma_0)$, and $\mathrm{len}(\gamma_1) = \mathrm{len}(\alpha_0)$. (This is the triangle inequality. Exercise!)

The parameterized version of the previous condition is that there exists δ' such that for any such triangle each of the pairs $\alpha_1,\ \beta_0;\ \beta_1,\ \gamma_0;\ \gamma_1,\ \alpha_0$ δ'-fellow travel, that is $d(\alpha_1(t), \beta_0(t)) \leq \delta'$ for $0 \leq t \leq \mathrm{len}(\alpha_1)$ and so on. The reader may wish to check that $\delta' = 2\delta$ suffices.



iii). We say that Γ has ***thin bigons*** if there is a δ such that for every pair of geodesics α and β with the same endpoints, $d(\alpha(t), \beta(t)) \leq \delta$ for $0 \leq t \leq \mathrm{len}(\alpha)$. (We call such a pair a geodesic bigon.) Warning: we are here treating Γ as a bona fide geodesic metric space. Accordingly, we must allow the endpoints of geodesics to occur in the interior of edges.

iv). We shall see that a group obeying any of these conditions is finitely presented. That is to say, there is a presentation

$$\langle \mathcal{G} \mid \mathcal{X} \rangle$$

where $\mathcal{X}$ consists of a finite set of words $w_1, \ldots, w_k$ on the set of generators $\mathcal{G}$ and G is isomorphic to the quotient of the free group on $\mathcal{G}$ by the normal closure of $\mathcal{X}$.

Another way of saying this is that a word w in the letters of $\mathcal{G}$ represents the identity in G if and only if it is freely equal to a product of conjugates of elements of $\mathcal{X}^{\pm 1}$. Thus a word in $\mathcal{G}$ represents the identity if is equal (in the free group on $\mathcal{X}$ to an expression of the form

$$\prod_{i=1}^{n} \rho_i w_{j_i}^{\pm 1} \rho_i^{-1}.$$

We say that G has a ***linear isoperimetric inequality*** if there is a constant K so that for any trivial word w we can satisfy this equation with $n \leq K\,\mathrm{len}(w)$.

v). We say that G has a ***sub-quadratic isoperimetric inequality*** if there is a sub-quadratic function $f(n)$ so that for any word w presenting the identity, w can be freely expressed as the product of at most $f(\mathrm{len}(w))$ conjugates of defining relators.

vi). We say that G has a ***Dehn's algorithm*** if there is a finite set of words $\mathcal{D} = \{v_1, \ldots, v_k\}$ with each v_i evaluating to the identity, so that if w is a word representing the identity, then there is some $v_i = u_0 u_1 \in \mathcal{D}$ so that $w = \alpha u_0 \beta$ and $\mathrm{len}(u_1) < \mathrm{len}(u_0)$.

We have seen that this gives an algorithm. Given a word w, we can replace it with the shorter word $\alpha u_1^{-1}\beta$ which evaluates to the same group element. If w does in fact represent the identity, after at most $\mathrm{len}(w)$ such moves, we are left with the empty word. If w does not represent the identity, after at most $\mathrm{len}(w)$ such moves, we are left with a word which we cannot shorten.

vii). We will say that ***geodesics diverge exponentially*** in Γ if there is a constant E and an exponential function e^{kt} ($k > 0$) with the following property: Suppose that α and β are geodesic rays based at a common point P. Suppose that there is a value R so that $d(\alpha(R), \beta(R)) \geq E$. Suppose now that p is a path connecting $\alpha(R+t)$ to $\beta(R+t)$ and that p lies outside the ball of radius $R+t$ around P. Then $\mathrm{len}(p) \geq e^{kt}$.

viii). We will say that ***geodesics diverge uniformly*** in Γ if there is a constant E and a function $f(t)$, $\lim_{t\to\infty} f(t) = \infty$, with the following property: Suppose that α and β are geodesic rays based at a common point P. Suppose that there is a value R so that $d(\alpha(R), \beta(R)) \geq E$. Suppose now that p is a path connecting $\alpha(R+t)$ to $\beta(R+t)$ and that p lies outside the ball of radius $R+t$ around P. Then $\mathrm{len}(p) \geq f(t)$.

Theorem. *All the above conditions are equivalent and independent of generating set. A group satisfying them is called* ***word hyperbolic****.*

Conditions i), ii), and vii) are equivalent to each other for any geodesic metric space and are characterizations of a type of hyperbolicity in such spaces called ***Gromov-Rips hyperbolicity***. In a class of spaces including complete simply connected riemannian manifolds they are also equivalent to linear isoperimetric inequality. In these spaces linear isoperimetric inequality means the existence of a constant K such that a closed loop l can always be spanned by a disk of area at most $k\,\mathrm{len}(l)$.

The thin bigons condition is equivalent to Gromov-Rips hyperbolicity in any graph with edges of unit length (see **[Pa2]**), but certainly not in arbitrary geodesic metric spaces (think of euclidean space!).

Conditions i), ii), iv), vi), and vii) can be found in general expositions such as **[ABC]**, **[B]**, **[C3]**, **[CDP]**, **[GH]**. Conditions v) and viii) can be found in **[Pa1]**.

1.2 Cayley graphs and group actions

We say that a metric space (X, d_X) is a ***geodesic metric space*** if distances in X are realized by geodesics in X That is, given $a, b \in X$, there is a path p connecting a and b so that $d_X(a,b) = \mathrm{len}(p)$ and $\mathrm{len}(p)$ is minimal among all paths connecting a and b.

A map $f : (X, d_X) \to (Y, d_Y)$ is a (λ, ϵ) ***quasi-isometric map*** if for all $a, b \in X$, $\frac{1}{\lambda} d_X(a,b) - \epsilon \leq d_Y(f(a), f(b)) \leq \lambda d_X(a,b) + \epsilon$. Note that such a map need only be almost continuous. A quasi-isometric map is a map which is (λ, ϵ) quasi-isometric for some (λ, ϵ). A quasi-isometric map is a ***quasi-isometry*** of X and Y if there is a quasi-isometric map g from Y to X and a constant k so that for all $x \in X$, $d_X(x, g(f(x))) \leq k$ and for all $y \in Y$, $d_X(x, f(g(y))) \leq k$. In this case we say X and Y are quasi-isometric. This is an equivalence relation on the class of metric spaces. Given a finitely generated group G all Cayley graphs of G are quasi-isometric.

Suppose now that (X, d) is a geodesic metric space. Suppose that G is a finitely generated group which acts by isometries on X. Suppose further that this action is discrete and co-compact. "Discrete" means that if $g_1, g_2, \ldots$ is a sequence of distinct group elements then for $x \in X$ the sequence $g_1 x, g_2 x, \ldots$ does not converge (this is slightly weaker than "proper discontinuity"). "Co-compact" means that the orbit space X/G is compact.

Let us fix a generating set $\mathcal{G}$ for G and pick a basepoint, $* \in X$. The map which takes $g \in G$ to $g(*) \in X$ takes 1 to $*$, is G-equivariant, and is finite-to-one since the action of G is discrete. For each $g \in \mathcal{G}$, we choose a path p_g from $*$ to $g(*)$. We now have a map

$$\phi : \Gamma_{\mathcal{G}}(G) \to X$$

defined by taking each vertex g of Γ to $g(*)$, and each edge (a, g, b) of Γ to $a(p_g)$. (Strictly speaking, we must parameterize both the edge (a, g, b) and the path p_g by the unit interval and take $(a, g, b)(t)$ to $a(p_g(t))$.)

Theorem (Milnor) [M]. *The map ϕ is a quasi-isometry.*

This is not a particularly difficult theorem — indeed, the reader may wish to try to prove it as an exercise. However, as mentioned in the Introduction, it was one of the turning points in the development of modern geometric group theory.

1.3 Groups acting on hyperbolic spaces.

In view of Milnor's Theorem, we will want to see that any space which is quasi-isometric to a hyperbolic space is itself hyperbolic. This will show that co-compact discrete action of a group G on a hyperbolic space X "transfers" the hyperbolicity of X to G. It will also show that hyperbolicity is independent of generating set.

2.1 Quasi-geodesics in a hyperbolic space

A ***geodesic*** γ is an isometric map of the interval $[0, \mathrm{len}(\gamma)]$. (You may take this a definition of $\mathrm{len}(\gamma)$.) A (λ, ϵ)-***quasi-geodesic*** α is a (λ, ϵ)-quasi-isometry of the interval $[0, \mathrm{len}(\alpha)]$. Here are several characterizations of the relationship between geodesics and quasi-geodesics in a hyperbolic space.

Given a path γ it is often convenient to extend it to a map defined on $[0, \infty)$ by setting $\gamma(t) = \gamma(\mathrm{len}(\gamma))$ for $t \geq \mathrm{len}(\gamma)$. We use this convention in the following theorem.

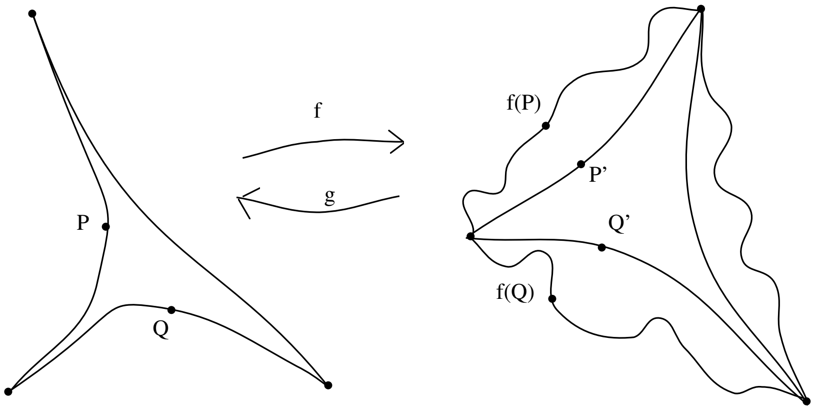
Theorem. *There are $N = N(\delta, \lambda, \epsilon)$, $K = K(\delta, \lambda, \epsilon)$, and $L = L(\delta, \lambda, \epsilon)$ so that if α is a (λ, ϵ)-quasi-geodesic and γ is a geodesic in X and X is δ-hyperbolic, and α and γ have the same endpoints, then each of the following hold:*
i). *Each of α and γ is contained in an N-neighbourhood of the other.*
ii). *α and γ asynchronously K-fellow travel. That is, there is a monotone surjective reparameterization $t \mapsto t'$ of $[0, \infty)$ so that for all t we have $d(\alpha(t), \gamma(t')) \leq N$.*
iii). *α progresses at some minimum rate along γ. That is, the reparameterization of ii) can be chosen so that $t \geq t_0 + L$ implies $t' \geq t_0' + 1$.*

This last is sometimes referred to as “progression in geodesic corridors."

2.2 Hyperbolicity is a quasi-isometry invariant

We now prove that hyperbolicity is a quasi-isometry invariant using the previous theorem and a picture.

Theorem. *Suppose that X and Y are quasi-isometric geodesic metric spaces. If Y is hyperbolic, so is X.*



2.3 Word and conjugacy problem for hyperbolic groups

It follows immediately from the existence of a Dehn's algorithm that a hyperbolic group has a highly efficient solution to its word problem. Most early solutions to the conjugacy problem used the boundary of a hyperbolic group, which will be described later. We will give a proof in §5.7 due to Gersten and Short which works in the more general setting of biautomatic groups, and which does not use the boundary.

2.4 Torsion in hyperbolic groups

There is a charming proof that hyperbolic groups have finitely many conjugacy classes of torsion elements based on the Dehn's algorithm. See for example, **[ABC]**.

2.5 The Rips Complex

A fundamentally important property of a hyperbolic group G is:

Theorem (Rips). *Let G be a hyperbolic group. Then G acts properly discontinuously on a finite dimensional contractible complex Y with compact quotient.*

Such a complex is easy to describe. We start with a Cayley graph Γ for G and take the geometric realization $Y = |P_d(\Gamma)|$ of the following abstract simplicial complex

$P_d(\Gamma)$ for an appropriate bound d. Vertex set of $P_d(\Gamma)$ is G and the simplices consist of subsets of G of diameter at most d in the word metric. It turns out that $d = 4\delta + 1$ always suffices. A very efficient proof of Rips' theorem is given in **[ABC]**.

This theorem has important homological implications for G. For example, if G is virtually torsion free (i.e., has a torsion free subgroup of finite index) then it has finite virtual cohomological dimension and in any case its rational homology and cohomology is finite dimensional. It is an important open problem whether a hyperbolic group is always virtually torsion free.

Another interesting consequence is that non-vanishing homology implies a lower bound for the hyperbolicity constant $\delta = \delta(\mathcal{G})$ for G among all finite generating sets for G.

2.6 The boundary of a hyperbolic group

For a hyperbolic group G (in fact more generally for any Gromov/Rips hyperbolic metric space) there is a natural compactification of G by adding a "boundary at infinity." Roughly speaking the boundary consists of the set of all ways to travel off to infinity. One way of making this precise is to define a ***geodesic ray*** in a Cayley graph Γ for G as an isometry of $[0, \infty)$ into Γ and to say two rays are ***equivalent*** if they fellow travel. The boundary ∂G (which in fact is a quasi-isometry invariant and thus does not depend on the choice of generating set) is the set of equivalence classes of geodesic rays. The topology on ∂G can be defined in a multitude of ways. Basicly two points are close if their rays fellow travel for a long time. One way to formalize this is to take the compact open topology on the set of rays and take the quotient topology on ∂G.

There are also many ways of describing how to attach this boundary to G or its Cayley graph Γ. One obtains a compact Hausdorff space into which both G and ∂G embed. We leave as an exercise how to attach the boundary: the construction is highly stable in the sense that any reasonable answer you give will be correct (see **[NS1]**).

3.1 Area of a word that represents the identity

As we have seen, a group G with presentation

$$\langle \mathcal{G} \mid \mathcal{X} \rangle$$

is isomorphic to $F_{\mathcal{G}}/N(\mathcal{X})$, where $F_{\mathcal{G}}$ is the free group on $\mathcal{G}$ and $N(\mathcal{X})$ is the normal closure of $\mathcal{X}$ in $F_{\mathcal{G}}$. It follows that a word $w \in \mathcal{G}^*$ represents the identity in G if and only if it is freely equal (that is, equal in $F_{\mathcal{G}}$) to an expression of the form

$$\prod_{i=1}^{k} \rho_i R_i \rho_i^{-1}$$

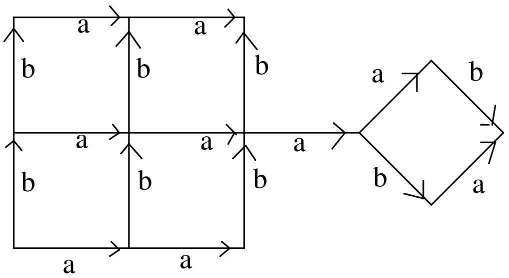
where each $R_i \in \mathcal{X}^{\pm 1}$. Thus, solving the word problem in G means determining the existence or non-existence of such an expression for each w.

A naive approach would be to start enumerating all such expressions in hopes of finding one freely equal to our given w. If w represents the identity, we must eventually find the expression that proves this. The problem with this approach is knowing when to give up if w does *not* represent the identity. Let us formalize this quandary.

Suppose we take a word w representing the identity in G. We define the ***area***, $A(w)$ to be the minimum k in any such expression for w.

3.2 Van Kampen diagrams

The choice of the word area is motivated the notion of a ***Van Kampen diagram*** for w. Such a diagram Δ is a labeled, simply connected sub-complex of the plane. Each edge of Δ is oriented and labeled by an element of $\mathcal{G}$. Reading the labels on the boundary of each 2-cell of Δ gives an element of $\mathcal{X}^{\pm 1}$. Δ is a Van Kampen diagram for w if reading the labels around the boundary of Δ gives w.



Van Kampen diagram for the word $a^2babab^{-1}a^{-2}ba^{-2}b^{-2}$ **in** $\langle a, b | a^{-1}b^{-1}ab\rangle$

Theorem. *Each Van Kampen diagram with k 2-cells for w gives a way of expressing w as a product $\prod_{i=1}^{k} \rho_i R_i \rho_i^{-1}$. Each product $\prod_{i=1}^{k} \rho_i R_i \rho_i^{-1}$ for w gives a Van Kampen diagram for w with at most k 2-cells. In particular, w has a Van Kampen diagram if and only if w represents the identity.*

Thus, we could have defined $A(w)$ to be the minimum number of 2-cells in a Van Kampen diagram for w.

3.3 Dehn's function of a presentation

We define the ***Dehn's function*** of the presentation $\langle \mathcal{G} \mid \mathcal{X}\rangle$ to be

$$\delta(n) = \max\{A(w) \mid \mathrm{len}(w) \leq n\}.$$

This maximum is taken over words presenting the identity. We say f is an ***isoperimetric function*** for $\langle \mathcal{G} \mid \mathcal{X}\rangle$ if $f(n) \geq \delta(n)$. (As we shall see below, "isoperimetric function" is often used in a sense between these two.)

3.4 Isoperimetric functions and the word problem

Either of these functions answers the question, "When should we give up?"

Theorem. $G = \langle \mathcal{G} \mid \mathcal{X} \rangle$ *has a soluble word problem if and only if it has a recursive Dehn's function if and only if it has a sub-recursive Dehn's function (and the same for isoperimetric functions).*

A function f is recursive if it can be computed by some computer program. It is sub-recursive if there is a recursive function g so that $f(n) \leq g(n)$ for all $n \geq 0$.

Proof. If a given word w represents the identity, we can eventually exhibit a Van Kampen diagram for it by sheer perseverance. The Dehn's function (if we can compute it or an upper bound for it) tells us the maximum number of 2-cells in any Van Kampen diagram we must consider. Now, the number of possible Van Kampen diagrams with a given number of two cells is unbounded, because such diagrams may have long 1-dimensional portions. However, the length of the boundary of such a diagram is at least twice the total length of the 1-dimensional portions. Thus any possible diagram for w has at most $\delta(\operatorname{len}(w))$ 2-cells and at most $\frac{1}{2}\operatorname{len}(w)$ 1-cells that are not on the boundary of some 2-cell. If we examine all of these and find none whose boundary is w, we know that w does not represent the identity. $\square$

One interpretation of this theorem is that there is only one way for a group to fail to have a soluble word problem: its Dehn function grows too fast.

3.5 Equivalence of isoperimetric functions

Actual Dehn functions and isoperimetric functions are too specific for our purposes, but there is a natural equivalence relation and partial order which are appropriate here. For positive valued functions on the natural numbers we say

$$f \preceq g$$

if there are positive A, B, C, D and E so that

$$f(n) \leq Ag(Bn + C) + Dn + E$$

for all n. We say that f and g are ***equivalent*** if $f \preceq g \preceq f$. Under this relation any two polynomials of the same positive degree are equivalent, and the equivalence class of linear and sub-linear functions is the least equivalence class. Any two exponentials (with base greater than 1) are equivalent.

Theorem. *Suppose G and H are quasi-isometric finitely presented groups. Then the Dehn's functions of G and H are equivalent. In particular, up to equivalence the Dehn's function of G is independent of presentation.*

One says G ***has a linear (quadratic, exponential, etc.) isoperimetric inequality***, meaning that what we have called its Dehn function is of the appropriate class.

3.6 The word problem is geometric

Theorem. *The property of having soluble word problem is a quasi-isometry invariant of finitely presented groups.*

This follows because the property of being sub-recursive is an equivalence class invariant of functions.

Since groups with unsolvable word problem exist, isoperimetric functions can be of enormously rapid growth. In fact, by their definition, non-sub-recursive functions can be thought to have "inconceivably rapid growth."

It is not hard to find groups with solvable word problem with very fast growing Dehn functions — for example, Gersten has pointed out **[G]** that for any k the function $f_k(n) = 2^{2^{\cdot^{\cdot^{\cdot^{n}}}}}$ (k levels of exponent) is a lower bound for the isoperimetric function for the group $\langle x, y | x^{x^y} = x^2 \rangle$ (the notation a^b is a shorthand for $b^{-1}ab$). Clearly, the algorithm proposed above for the word problem — enumerating all Van Kampen diagrams up to the size given by the Dehn function — is absurd when one has a Dehn function that grows this fast (or even exponentially fast). For specific groups much faster algorithms can often be found.

4 JSJ decomposition.

The name JSJ refers to Johannson, and Jaco and Shalen, who developed a theory, building on earlier ideas of Waldhausen, for cutting irreducible three-dimensional manifolds into pieces along tori and annuli. Later Thurston explained these decompositions from a geometric point of view in his famous "Geometrization Conjecture" for 3-manifolds. One can describe JSJ decomposition in terms of amalgamated product and HNN decomposition of the relevant fundamental group, and in fact a purely group theoretic version of the theory has been worked out by Kropholler and Roller.

Around 1992 Zlil Sela pointed out that analogous decompositions appear to exist for groups in a much broader range of situations. First Sela did this for torsion free hyperbolic groups and then Sela and Rips extended it to general torsion free finitely presented groups. This is currently a very active area of research, and we can only touch on some of its coarsest elements here. Since some of the major players are here at this workshop (Swarup, Bowditch), we can hope to learn more in seminars. In fact we are indebted to them for the information in this section, although any errors in it are our responsibility.

The theory also has origins in work of Stallings on "ends of groups." The ***set of ends*** of a locally compact space X is the limit over larger and larger compact subsets K of X of the set of components of $X - K$. The set of ends of a finitely generated group G is the set of ends of any connected space on which G acts freely with compact quotient — for example one may take a Cayley graph of the group. The definition turns out to be independent of choices. For example, $\mathbb{Z}$ has two ends, $\mathbb{Z}^n$ has one end for $n > 1$, and a free group on two or more generators has infinitely many ends, since its Cayley graph is an infinitely branching tree. For a hyperbolic group it is known that the set of ends coincides with the set of components of the boundary of the group.

Stallings showed that a finitely generated group can have only 0 (if the group is finite), 1, 2, or infinitely many ends. Moreover, if 2 ends then the group is virtually infinite cyclic and if infinitely many then the group can be "split" along a finite subgroup, that is, it is an amalgamated free product or HNN extension amalgamated along a finite subgroup. In the latter case of "decomposing G along a finite subgroup," one might try to iterate the decomposition if the component groups that are being amalgamated still have infinitely many ends. Work of Dunwoody **[Du]** shows that for a finitely presented group this iteration eventually ends. One codes the result in what is called a "graph of groups." This is a finite graph with groups assigned to vertices of the graph and subgroups of the vertex groups assigned to adjacent edges of the graph as a scheme for describing how to repeatedly amalgamate the vertex groups along the edge groups using amalgamated free products or HNN extensions. In our case all the edge groups are finite and the vertex groups each have at most one end, since they cannot be further decomposed. This is, as it were, the first stage of JSJ decomposition, and leaves us with one-ended groups to decompose.

We now restrict to one-ended hyperbolic groups. In this case JSJ decomposition concerns splitting along virtually infinite cyclic subgroups. The main question is when such splittings exist. We have:

Theorem (Paulin and Rips). *If G is a one ended hyperbolic group with $|Out(G)| = \infty$ then G splits along a virtually infinite cyclic group.*

By $Out(G)$ we mean the outer automorphism group: the quotient of $Aut(G)$ by the group $Inn(G)$ of inner automorphisms, that is automorphisms induced by conjugation by a group element.

Theorem (Swarup and Scott, Bowditch). *If G is a hyperbolic group with one end and H is a virtually infinite cyclic subgroup with the relative number of ends $e(G,H) \geq 2$ then G splits along some virtually infinite cyclic subgroup (except for some virtual surface groups).*

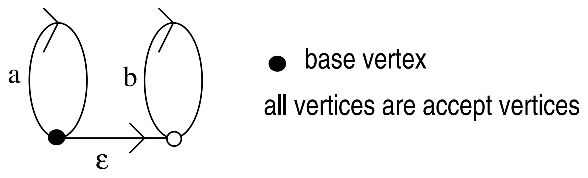
Swarup and Scott proved the torsion free case. Bowditch's proof is very different and uses the boundary. It also shows that the existence of such a splitting is a geometric property, i.e., invariant under quasi-isometry.

The number of relative ends is given by taking a connected space X on which G acts freely and cocompactly as before and counting the ends of X/H. One must exclude virtual surface groups in the above theorem, since they include triangle groups which do not split despite having cyclic subgroups with $e(G,H) = 2$.

Rips and Sela have shown that the JSJ splittings are unique up to an appropriate equivalence relation in special cases.

5.1 Regular languages, finite state automata, and automatic structures

A ***finite state automaton*** $\mathcal{M}$ over an alphabet $\mathcal{A}$ is a finite directed, labeled graph equipped with a base vertex and a set of preferred vertices or "accept vertices." The edges of $\mathcal{M}$ are labeled by letters of $\mathcal{A} \cup \{\epsilon\}$, where ϵ is an additional symbol that stands for "empty". A finite state automaton M defines a language $L(\mathcal{M}) \subset \mathcal{A}^*$. $L(\mathcal{M})$ is the set of words labelling paths in $\mathcal{M}$ starting from the base vertex of $\mathcal{M}$ and ending at accept vertices of $\mathcal{M}$. For example, the language defined by the following finite state automaton is $\{a^n b^m : n, m \geq 0\}$.



A language is a ***regular*** language if it is the language of some finite state automaton. One also speaks of the ***language accepted by*** the finite state automaton.

We say the finite state automaton $\mathcal{M}$ is ***deterministic*** if no ϵ-edges occur and each vertex of M has exactly one edge emanating from it for each element of $\mathcal{A}$.

A finite state automaton is a model of a simple computing device. This is best seen in the deterministic case. A deterministic finite state automaton can be written as a 5-tuple. $\mathcal{M} = (\mathcal{A}, S, \tau, s_0, A)$, where $\mathcal{A}$ is the alphabet, S is the set of vertices, or ***states***, $\tau : S \times \mathcal{A} \to S$ is the ***transition function*** given by the edges of $\mathcal{M}$, s_0 is the base vertex or ***start state*** and $A \subset S$ is the set of preferred vertices or ***accept states***.

We imagine that $\mathcal{M}$ performs computations on words to determine whether or not they lie in $L(\mathcal{M})$. Given the word $w = a_1 \dots a_k$, the computation proceeds as follows: $\mathcal{M}$ starts in the state $s(0) = s_0$ and reads a_1. This causes it to enter state $s(1) = \tau(s_0, a_1)$. It proceeds in this way reading the letters of w and changing states according to the formula $s(j) = \tau(s(j-1), a_j)$. It concludes that $w \in L(\mathcal{M})$ if and only if $s(k) \in A$.

There are many characterizations known for regular languages. The following theorem, which gives some of them, is a worthwhile exercise. First a definition that we will return to later. We define the ***cone*** of a word w with respect to a language L. It is the set

$$C_w(L) := \{v \in \mathcal{A}^* : wv \in L\}.$$

Theorem/Exercise. *The following are equivalent for a language* $L \subset \mathcal{A}^*$.
i). L *is a regular language (i.e., accepted by a finite state automaton);*
ii). *There are only finitely many different cones* $C_w(L)$ *as* w *runs through* $\mathcal{A}^*$*;*
iii). L *is accepted by some deterministic finite state automaton.*

(Hint: to show ii) ⇒ iii) use the set of cones as the states of the finite state automaton.)

□

Another standard way to characterize regular languages is in terms of closure properties. We will not discuss this characterization, but it is a good exercise to prove the closure properties.

Theorem. *If* $L_1, L_2 \subset \mathcal{A}^*$ *are regular languages then so are:* $L_1 \cup L_2$, $L_1 \cap L_2$, $L_1L_2 := \{w_1w_2 : w_1 \in L_1, w_2 \in L_2\}$, $\mathcal{A}^* - L_1$, $L_1^* := \{w : \exists n \geq 0$ and $w_1, \dots, w_n \in L_1$ with $w = w_1 \dots w_n\}$. $\square$

In addition to using finite state automaton to examine single words $w \in \mathcal{A}^*$, we will need to use them to examine pairs $(w, w') \in \mathcal{A}^* \times \mathcal{A}^*$. In order to do this, we use a technical trick: We form the ***padded product alphabet***

$$\mathcal{A}^2 = (\mathcal{A} \times \mathcal{A}) \bigcup (\mathcal{A} \times \{\$\}) \bigcup (\{\$\} \times \mathcal{A})$$

where $\$$ is a symbol not in $\mathcal{A}$. We then embed $\mathcal{A}^* \times \mathcal{A}^*$ into $(\mathcal{A}^2)^*$ by writing $(a_1 \dots a_j, b_1 \dots b_k)$ as

$(a_1, b_1) \dots (a_j, b_j)$, if $k = j$,

$(a_1, b_1) \dots (a_j, b_j)(\$, b_{j+1}) \dots (\$, b_k)$, if $j < k$, and as

$(a_1, b_1) \dots (a_k, b_k)(a_{k+1}, \$) \dots (a_j, \$)$ if $j > k$.

When we speak of a subset of $\mathcal{A}^* \times \mathcal{A}^*$ being the language of a finite state automaton we mean it with respect to the padded product alphabet as above. A finite state automaton for such a language is often called a ***(synchronous) two-tape finite state automaton***, since one imagines it being fed the two input words on two separate "input tapes." If we extend the alphabet to include also pairs (a, ϵ) and (ϵ, a), where ϵ denotes the empty word, one obtains what is called an ***asynchronous two-tape automaton***, since the automaton can read a letter on only one tape at a time and thus read the two "tapes" asynchronously.

Another useful closure property of regular languages (again an exercise to prove) is:

Proposition. *If* $L \subset \mathcal{A}^* \times \mathcal{A}^*$ *is the language of a (possibly asynchronous) two-tape automaton, then its projection onto the first factor* $\{w : \exists v \ (w, v) \in L\}$ *is a regular language.* $\square$

We now give the original automaton-theoretic definition of an automatic structure for G. The characterization of the first theorem of §5.2 gives a more useful working definition and is often used nowadays.

An ***automatic structure*** for G consists of the following:

1) A finite set $\mathcal{G}$ together with a map $\mathcal{G} \to G$. We write this map $g \mapsto \overline{g}$ and extend it to $\mathcal{G}^*$ as a monoid homomorphism.
2) A regular language $L \subset \mathcal{G}^*$ so that $\overline{L} = G$.
3) A synchronous two-tape automaton $\mathcal{M}_0$ so that

$$L(\mathcal{M}_0) = \{(w, w') \in L \times L \mid \overline{w} = \overline{w'}\}.$$

4) For each $a \in \mathcal{G}$ a synchronous two-tape automaton $\mathcal{M}_a$ so that

$$L(\mathcal{M}_a) = \{(w, w') \in L \times L \mid \overline{w} = \overline{w'a}\}.$$

Such a structure fulfils our desire to build the Cayley graph "on the cheap". The language L gives us names for the vertices of $\Gamma_{\mathcal{G}}$, and since it is regular, it is cheap to determine when we have such a name. $\mathcal{M}_0$ cheaply determines when two names name the same vertex, and $\{\mathcal{M}_a\}$ cheaply determines when two such names name an edge. All this can be made even cheaper by proving:

Theorem. *Any automatic structure L for G contains a sub-language which is an automatic structure which bijects to G.*

For a proof of this and other basics of automatic structures see **[ECHLPT]** or **[BGSS]**.

5.2 Fellow traveler property

There is a geometric condition which tells us when a regular language which surjects to the group is an automatic structure. We say that L has the ***fellow traveler property*** if there exists K such that for any pair of words $w, w' \in L$ so that w and w' end at most an edge apart, $d(w(t), w'(t)) \leq K$ for all $t \in [0, \infty)$.

Theorem. *Suppose L is a regular language that surjects to G. Then L is an automatic structure if and only if L has the fellow traveler property.*

Proof. We first show that an automatic structure has the fellow-traveler property. This uses the following simple but basic lemma about regular languages.

Lemma. *Let $L \subset \mathcal{G}^*$ be a regular language. Then there exists a bound s such that if w is an initial segment of an L-word, that is, there exists a v with $wv \in L$, then there exists such a v of length at most s.*

Proof. Let s be the number of states in an automaton for L. When we feed w to the automaton, the fact that it can be extended to an L-word means it ends at a state from which an accept state can be reached. This accept state can then clearly be reached in at most s steps (in fact $s - 1$ suffices). □

Applying this lemma to the language accepted by the comparator automaton shows that a pair of words being compared are always at most s steps away from being 1 letter apart, so their values are at most $2s + 1$ apart, proving the fellow traveler property.

The fact that the fellow-traveler property implies the existence of comparator automata follows easily from the following proposition, and is left to the reader. □

Proposition. *Let G be a group with finite generating set $\mathcal{G}$. Then for any $K > 0$ and any $g \in G$ the languages*

$$M_K = \{(w, w') \in \mathcal{G}^* \times \mathcal{G}^* : w \text{ and } w' \; K\text{-fellow-travel}\}$$
$$M_K(g) = \{(w, w') \in M_K : \overline{w'} = \overline{w}g\}$$

are regular languages (actually, to be precise, we must use the corresponding padded languages, as described above).

Proof. Denote $g(t) = w(t)^{-1}w'(t)$ for each t. Note that if we know $g(t)$ and know the next letters a and b of w and w' respectively then we can compute $g(t+1)$, since $g(t+1) = (w(t)a)^{-1}w'(t)b = a^{-1}g(t)b$. The words w and w' fellow-travel if and only if $g(t) \in B(K)$ for all t. We can thus build a finite state automaton to recognise the languages M_K and $M_K(g)$ by using the ball $B(K)$ as set of states plus an additional "fail state" and letting the (a, b)-edge from state $g \in B(K)$ lead to $a^{-1}gb$ if this element is in $B(K)$ and to the fail state otherwise. All edges from the fail state lead back to the fail state and the reader can easily work out what the edges labelled by $(a,\$)$ and $(\$,b)$ should do, where $\$$ is the pad character. Start state is 1 and accept state is g for the language $M_K(g)$ (if $g \in B(K)$) and is the whole of $B(K)$ for the language M_K. $\square$

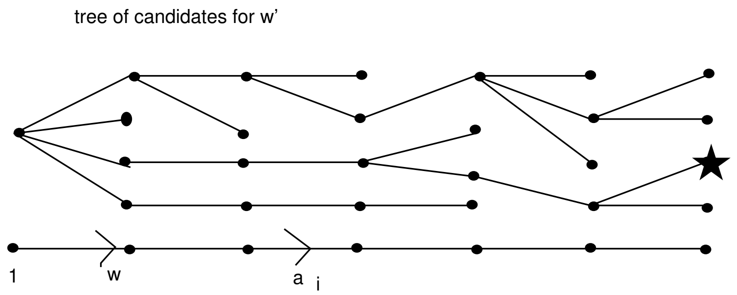
5.3 Quadratic isoperimetric equality and quadratic time word problem

From the characterization of automatic structures by fellow-traveler property one deduces:

Corollary. *If L is an automatic structure which bijects to G, then there are λ and ϵ so that L consists of (λ,ϵ)-quasi-geodesics.*

With a little more work one has

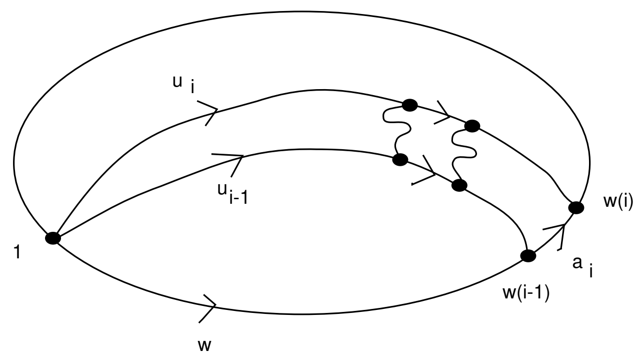
Theorem. *Suppose L is an automatic structure which bijects to G. Then there is an algorithm which takes as input the letter $a \in \mathcal{G}$ and $w \in L$ and finds the word in L representing $\overline{wa}$ This is accomplished in at most $k\,\mathrm{len}(w)$ steps for some k that is independent of w.*



As we walk the word w through the comparator machine for a we produce a tree of candidates for w'. Since $w'(t)$ is always within distance K of $w(t)$ at each stage there are at most $|B(K)|$ candidates for $w'(t)$ where $B(k)$ is the ball of radius K. Since L bijects to G, two different candidates for w' that have reached the same value $w'(t)$ must have reached different states of the comparator automaton. Thus the number of candidates for w' at time t is at most $|B(K)||S|$, where S is the set of states of the comparator automaton. Thus the tree of possibilities remains of bounded width. When we reach the end of w the comparator machine declares the winning w'. $\square$

From these and the following picture we deduce

Theorem. *If G is automatic, then*
1) G is finitely presented.
2) G has at most quadratic isoperimetric inequality.
3) G has quadratic time word problem.



5.4 Closure properties

Theorem. *The set of automatic groups is closed under the following:*

i) Free products.
ii) Direct products.
iii) Free factors.
iv) Finite index subgroups.
v) Finite index supergroups.
vi) HNN extensions and free products with amalgamation over finite subgroups.

The set of bi-automatic groups (see below) is closed under all the above except possibly finite index supergroups and additionally is closed under central quotients and direct factors.

Again, we direct the reader to **[BGSS]** and **[ECHLPT]**. The result on biautomatic groups is due to Lee Mosher, **[Mo2]**.

As we shall see below, the sets of automatic, biautomatic, hyperbolic, groups are closed under passing to "rational" subgroups.

5.5 Famous classes of automatic groups.

Theorem. *Automatic groups include the following:*

i) Hyperbolic groups including: free groups, finite groups, most small cancellation groups, fundamental groups of closed negatively curved manifolds and other negatively curved spaces **[ECHLPT]**
ii) Small cancellation groups **[GS1]**.
iii) Fundamental groups of three manifolds which satisfy Thurston's geometrization conjecture, except for those containing a nil or solv manifold as a connected sum component **[S2]**, **[ECHLPT]**.
iv) Coxeter groups **[BH]**.
v) Mapping class groups **[Mo1]**.
vi) Braid groups and more generally, Artin groups of finite type **[Ch]**.
vii) Central extensions of hyperbolic groups **[NR]**.
vii) Many amalgams of hyperbolic groups along rational subgroups, **[BGSS]** *and* **[S1]**.
ix) Many groups that act on affine buildings **[CS]**.

In fact i), iii), vi) and vii) are known to be bi-automatic (for definition of bi-automatic see below). We will not define the various classes of the above theorem. Some will be described in other lectures of this workshop. Suffice it to say that a lot of interesting groups are automatic.

5.6 Cone types and falsification by fellow-traveler

Regular languages are closely related to Cannon's notion of ***cone types***. We will consider paths in the Cayley graph $\Gamma_{\mathcal{G}}(G)$. Edge paths are given by translates of words. We will need to consider paths which are not necessarily edge paths. In fact, it will suffice to consider paths which begin at vertices and end either at vertices or midpoints of edges, and we may assume our paths are parametrized by arc length.

We say a path p is ***outbound*** if $d(1, p(t))$ is a strictly increasing function of t. For a give $g \in G$, the ***cone*** at g, denoted $C'(g)$ is the set of all outbound paths starting at g. Thus if (g, a, g') is an edge with $\mathrm{len}(g') = \mathrm{len}(g) + 1$, then $C'(g)$ contains the edge (g, a, g'). If, however (g, a, g') is an edge with $\mathrm{len}(g') = \mathrm{len}(g)$, then $C'(g)$ does not contain the edge (g, a, g'), but it does contain the path consisting of the first half of that edge. We define the ***cone type*** of g, denoted $C(g)$ to be $g^{-1}C'(g)$.

We say $\Gamma_{\mathcal{G}}(G)$ has the ***falsification by fellow traveler property*** if there is a constant K so that if $w \in \mathcal{G}^*$ is not geodesic, then there is $w' \in \mathcal{G}^*$ so that

- $\overline{w} = \overline{w'}$,
- $\mathrm{len}(w') < \mathrm{len}(w)$, and
- w and w', K-fellow travel, that is $d(w(t), w'(t)) \leq K$ for all t.

Theorem. *If $\Gamma_{\mathcal{G}}(G)$ has the falsification by fellow traveler property, then $\Gamma_{\mathcal{G}}(G)$ has finitely many cone types.*

Proof. Suppose K is the constant for the falsification by fellow traveler property. Given $g \in G$ we define a function f_g on $B(K)$ by $f_g(h) = d(1, gh) - d(1, g)$, that is the relative distance of gh from 1 compared to that of g. The falsification by fellow traveler property implies that a path w from g is outbound if and only if there is no path w' from a point gh to the endpoint $g\overline{w}$ with $h \in B(K)$ and $\mathrm{len}(w') + d(1, gh) < \mathrm{len}(w) + d(1, g)$. This inequality can be written $\mathrm{len}(w') < \mathrm{len}(w) - f_g(h)$, so it follows that the cone type at g is determined by the function f_g. Since there can only be finitely many such functions (note that $|f_g|$ is bounded by K) there are only finitely many cone types. □

Using the first theorem of section 5.1 we deduce:

Theorem. *If $\Gamma_{\mathcal{G}}(G)$ has the falsification by fellow traveler property, then the set of geodesics in $\mathcal{G}^*$ is a regular language.*

Proof. The language of geodesics will have finitely many cones in the sense discussed earlier. □

The above results have their roots in **[C1]** and can be read in **[NS2]**.

Because of the issue of half-edges in cone types, cone types are more sensitive objects than the "language cones" described earlier. Thus the converse to "finitely many cone types implies language of geodesics is regular" is not known though it is close to being true. With the proper definitions one has

Theorem. *Suppose the language of geodesics in the barycentric subdivision of $\Gamma_{\mathcal{G}}(G)$ is regular. Then $\Gamma_{\mathcal{G}}(G)$ has finitely many cone types.*

or alternately,

Theorem. *Suppose $\Gamma_{\mathcal{G}}(G)$ has no edges whose endpoints are equidistant from the identity. Then $\Gamma_{\mathcal{G}}(G)$ has finitely many cone types if and only if the language of geodesics in $\mathcal{G}^*$ is regular.*

5.7 Bi-automaticity

We have seen that a group is automatic if there is a regular normal form $L \subset \mathcal{G}^*$ and finite state automata are capable of discovering right multiplication. That is, for each $a \in \mathcal{G}$ there is a finite state automaton $\mathcal{M}_a$ which discovers $\{w, w' \in L^2 \mid \overline{w'} = \overline{wa}\}$. There is an important subclass of automatic groups, namely those possessing automatic structures in which left multiplication can also be discovered by finite state automata. Such groups are called ***bi-automatic***. Specifically, an automatic structure is a ***bi-automatic structure*** if for each $a \in \mathcal{G}$ there is a finite state automaton $\mathcal{M}_a$ which discovers $\{w, w' \in L^2 \mid \overline{w'} = \overline{aw}\}$. Another way of saying this is that both the language L and the language L^{-1} should be automatic structures.

One can also characterize bi-automaticity in terms of fellow-travelling:

Theorem. *If $\mathcal{G}$ is a finite generating set for G and $L \subset \mathcal{G}^*$ a regular language which surjects to G then L is a bi-automatic structure if it possesses the following fellow traveler property: There is a constant K so that if w and w' are L-words beginning and ending at most distance 1 apart in Γ (so $\overline{aw} = \overline{w'a'}$ with $a, a' \in \mathcal{G}$) then $d(\overline{a}w(t), w'(t)) \leq K$ for all t. (Recall that the path $\overline{a}w$ is the path labelled by w starting at $\overline{a}$.)*

Theorem (Hyperbolic groups are geodesically biautomatic). *If G is a hyperbolic group and $\mathcal{G}$ is any generating set then the language L of geodesic words in $\mathcal{G}$ is a biautomatic structure on G.*

Proof. The fellow-traveler property for geodesics is immediate from the definition of hyperbolic group. The falsification by fellow traveler property then follows: given a non-geodesic word, replace the shortest non-geodesic initial segment of it by a geodesic to get a shorter word which fellow travels it. By the results above, it follows that the language L of geodesics is a regular language. It is thus an automatic structure, and in fact biautomatic since $L = L^{-1}$. □

The following is an open problem.

Problem. *Is every automatic group biautomatic?*

It is certainly not true that every automatic structure is biautomatic, but it could be true that if a group has some automatic structure then some automatic structure on the group is biautomatic.

The above problem would be answered in the negative if one could find an automatic group with unsolvable conjugacy problem.

Theorem [GS]. *If G is bi-automatic then G has solvable conjugacy problem.*

Proof. One version of the proof relies on two-tape automata. Suppose that we are given two elements g and g'. We consider the set $L(g, g')$ of all possible pairs of normal form words, w and w' so that $g\overline{w} = \overline{w'}g$. Now biautomaticity ensures that all such pairs of paths fellow travel with fellow traveler constant $\delta = K \max\{\mathrm{len}(g), \mathrm{len}(g')\}$. In particular, we can use the same methods that we used in §5.2 to give an explicit construction of a finite state automaton $\mathcal{M}(g, g')$ which discovers the language $L(g, g')$. This automaton has size at most $\beta = |B(\delta)| + 1$. Notice that g and g' are conjugate if and only if there is some pair (w, w) in $L(g, g')$. That is, g and g' are conjugate if and only if $L(g, g') \cap \Delta \neq \emptyset$ where Δ denotes the diagonal in $\mathcal{G}^* \times \mathcal{G}^*$. Now Δ is clearly a regular language, so the intersection $L(g, g') \cap \Delta$ also is. Moreover, it is easy to see that β is also a bound on the size of an automaton for $L(g, g') \cap \Delta$. So to determine if $L(g, g') \cap \Delta$ is empty or not we need only check words of length at most β. □

It is worth noting that the above proof actually shows that the set $\{w \in L : \overline{w}^{-1} g \overline{w} = g'\}$ is a regular sublanguage of L. This is an important fact that we will return to in section 7.

5.8 Asynchronous automaticity

We say two paths w and w' ***asynchronously*** K***-fellow-travel*** if they can be made to K-fellow-travel by reparameterizing them; that is, there exist monotonic surjective reparameterizations $t \mapsto t'$ and $t \mapsto t''$ of $[0, \infty)$ such that $d(w(t'), w(t'')) \leq K$ for all t.

Recall that an automatic structure on G can be described as a regular normal form L on the group with the (synchronous) K-fellow-traveler property for some K. One definition of ***asynchronous automatic structure*** is a regular normal form with the asynchronous fellow traveler property. There is also a machine-theoretic definition by using asynchronous two-tape automata as comparator automata. These two definitions are not quite equivalent but are as good as equivalent: they are equivalent for finite-to-one normal forms, and one can always find a sublanguage of an asynchronous automatic structure (using either definition) which is not just a finite to one asynchronous automatic structure but even a bijective one.

Asynchronous automatic structures are generally much weaker than automatic structures. For example, we have already seen that automaticity implies quadratic isoperimetric inequality. For asynchronous automaticity we only have:

Theorem. *If G has an asynchronous automatic structure then it has exponentially bounded isoperimetric function.*

The Baumslag-Solitar groups $\langle x, y : y^{-1}x^p y = x^q \rangle$ with p, q positive integers are asynchronously automatic groups which have exponential isoperimetric functions when $p \neq q$ (**[BGSS]**).

6 Equivalence of automatic structures.

Let L and L' be two automatic structures on G, using generating sets $\mathcal{G}$ and $\mathcal{G}'$ respectively. Then we may consider $L \cup L'$ as a language on the alphabet $\mathcal{G} \cup \mathcal{G}'$. We say L is ***equivalent*** to L' if $L \cup L'$ is an asynchronous automatic structure. Equivalently,

- there is a K such that every L-word K-fellow-travels an L'-word with the same value and vice versa.

This definition is useful also for asynchronously automatic structures, bi-automatic structures, and even for rational structures (a ***rational structure*** on G is simply a regular normal form — that is a regular language $L \subset \mathcal{G}^*$ which surjects to G, where $\mathcal{G}$ is a finite generating set). In the case of rational structures we use the definition of $\bullet$.

This turns out to be the right definition of equivalence to make automatic (and biautomatic and asynchronously automatic) structures independent of generating set — to any such structure on G with respect to one finite generating set one can find an equivalent one with respect to any other finite generating set. The proof of this is not hard, though it needs a little bit of care in the synchronous case.

We can therefore avoid reference to the generating set and denote by $\mathfrak{A}(G)$ the set of automatic structures on G mod equivalence and $\mathrm{B}\mathfrak{A}(G)$ the set of bi-automatic structures on G mod equivalence. Abuse of notation lets us consider $\mathrm{B}\mathfrak{A}(G)$ as a subset of $\mathfrak{A}(G)$. We will use a subscript "async" for the asynchronous versions of these. Here is an example result.

Theorem. *If G is hyperbolic then $\mathfrak{A}(G)$ consists of a single point.*

Proof. We may assume our automatic structure is finite to one in which case we have pointed out earlier that it consists of quasi-geodesics and that quasi-geodesics fellow-travel geodesics in a hyperbolic group. Thus the automatic structure is equivalent to the geodesic automatic structure. $\square$

Much of the theory of "rational subgroups" is best stated in terms of this equivalence relation. We will return to the following theorem in §7, but it is worth stating here. Given an automatic structure L on G a subgroup H is ***rational*** if $\{w \in L : \overline{w} \in H\}$ is a regular sublanguage of L.

Theorem. i) *Suppose H is a subgroup on G and L and L' are automatic structures on G. Then H is L-rational if and only if H is L' rational. If H is L-rational, then $[L]$ induces a unique automatic structure up to equivalence on H. Consequently,*

ii) *Suppose H is a subgroup of G which is rational in every automatic (resp. bi-automatic) structure on G. Then there is a map $\mathfrak{A}(G) \to \mathfrak{A}(H)$ (resp. $\mathrm{B}\mathfrak{A}(G) \to \mathrm{B}\mathfrak{A}(H)$).*

iii) *Suppose H is finite index in G. Then, up to equivalence, every automatic structure on H induces an automatic structure on G. Combining this with ii) gives a bijection between $\mathfrak{A}(H)$ and $\mathfrak{A}(G)$.*
iv) *The above are also valid in the asynchronous case.*

Complete information about $\mathfrak{A}(G)$ or $\mathfrak{A}_{async}(G)$ is usually hard to come by, although they have been determined for some groups (see e.g., **[NS1]**). For example, for a hyperbolic group we have already proved that all automatic structures are equivalent to each other. This is also true of asynchronous automatic structures for free groups and surface groups (in the surface case this is a result of Brady), but probably not for fundamental groups of hyperbolic 3-manifolds. In fact, if some cover of the hyperbolic 3-manifold fibers over the circle — and Thurston has conjectured that this always happens — then one can construct enormous numbers of inequivalent asynchronously automatic structures on its fundamental group. It would therefore be very exciting to find a hyperbolic 3-manifold group with few such structures! — an open problem.

Here is one theorem from **[NS1]**.

Theorem. **i)** *Let G be virtually $\mathbb{Z}^n$. Then $\mathfrak{A}(G)$ naturally bijects to the set of ordered rational linear triangulations of S^{n-1}.*
ii) *Let $[L_T]$ be the class corresponding to triangulation T. Then there is a bi-automatic structure in $[L_T]$ if and only if T is invariant under the action of the finite group $G/\mathbb{Z}^n$ on T.*
iii) *Let H be a subgroup of G. Then H is $[L_T]$-rational if and only if the great subsphere corresponding to H is a sub-complex of T.*

The equivalence relation on automatic structures leads naturally to a version of bi-automaticity. There is an action of G on $\mathfrak{A}(G)$. One way to describe this action is that given an element g and a language $[L]$ we define the class $g[L]$ to be those automatic structures which fellow travel the collection of translated paths gL. In fact, we could simply pick a word u for g and take $g[L]$ to be $[uL]$. (The language uL is formed by prefixing u to each word of L.) An automatic structure L is "asynchronous-synchronous bi-automatic" if and only if $[L]$ is fixed under the action of G. We leave the definition of "asynchronous-synchronous bi-automatic" as an exercise. Although this property is slightly less than full bi-automaticity, it is sufficient to solve the conjugacy problem and do the other nice things that follow from bi-automaticity that are discussed below, but this takes some work to prove.

7.1 Rational and quasiconvex subsets

The theory of rational subgroups originated with Gersten and Short **[GS]**.

Let L be a rational structure on G (recall that this just means L is a regular language that surjects to G — no fellow traveler property need be assumed — however our real interest is automatic structures). A subset $S \subset G$ is ***L-rational*** if $\{w \in L : \overline{w} \in S\}$ is a regular language. A subset $S \subset G$ is ***L-quasiconvex*** if there is a bound K such that if $w \in L$ satisfies $\overline{w} \in S$ then $w(t)$ is never distance more than K from S.

Proposition. 1. *Every L-rational subset is L-quasiconvex. The converse holds for subgroups.*

2. *Let L and L' be equivalent rational structures for G. Then $S \subset G$ is L-rational (resp. L-quasiconvex) if and only if it is L'-rational (resp. L'-quasiconvex).*

Proof First part. The quasiconvexity of a rational subset follows quickly from the lemma in §5.2.

Now suppose S is a quasiconvex subgroup. Then if $w = a_1 \dots a_n \in L$ evaluates into S we can find $1 = g_0, g_1, \dots, g_{n-1}, g_n = 1 \in B(K)$ such that $\overline{a_1 \dots a_j} g_j \in S$ for each j. Since $\overline{w} = \prod g_i^{-1} a_i g_i$ we see that the set $\mathcal{A} := \{g^-1\overline{a}g \in S : g \in B(K)$ and $a \in \mathcal{G}\}$ generates S. This is clearly a finite set, so S is finitely generated.

We now consider $\{(w, w') : w \in L, w' \in \mathcal{A}^*, w$ and w' K-fellow travel$\}$. By the method of §5.2 and the fact that intersection of regular languages is regular, this is the language of an asynchronous two-tape automaton. Thus the image of projection onto the first factor is regular. But this is $\{w \in L : \overline{w} \in S\}$, so S is rational.

Second part. For quasiconvexity the proposition is trivial. Assume S is L-rational, so $N = \{w \in L : \overline{w} \in S\}$ is a regular language. Let $N' = \{w \in L' : \overline{w} \in S\}$. By the argument of §5.2 there is an asynchronous two-tape automaton $\mathcal{T}$ whose language is $\{(u, v) \in \mathcal{G}^* \times (\mathcal{G}')^* : \overline{u} = \overline{v},\ u$ and v asynchronously K-fellow travel $\}$. Since intersection of regular languages is regular, $\{(u, v) \in N \times L' : \overline{u} = \overline{v},\ u$ and v asynchronously K-fellow travel $\}$ is a non-deterministic asynchronous two-tape language. Its projection onto the second factor is N', and is hence regular as required. □

7.2 Rational subgroups inherit automatic or hyperbolic structures

The fact that a rational subgroup of an automatic group inherits an automatic structure has already been said in §6. It is now easy to prove. We just look at the two-tape language described in the above "Proof First part" of §7.1 and project it onto its second factor instead of its first. The same proof applies to asynchronously automatic and to biautomatic. □

The analogous result for hyperbolicity is also true and is worth further discussion. For a hyperbolic group the rationality concept for subgroups and subsets does not depend on the choice of automatic structure, since there is just the one automatic structure up to equivalence. For a subgroup H, rationality is, as we have just seen, equivalent to quasiconvexity with respect to geodesics. It follows quite easily that H-geodesics are quasi-geodesic in G and therefore fellow travel the corresponding G-geodesics. The thin triangles condition for H now follows, so we have shown:

Theorem. *Rational subgroups of hyperbolic groups are hyperbolic.* □

It is not true that any hyperbolic subgroup of a hyperbolic group is rational. We will return to this in a moment. However:

Exercise. *Any finitely generated subgroup of a finitely generated free group is rational.*

Since it is clear that intersection of two rational subgroups is rational (since the intersection of two regular languages is regular), we obtain the corollary:

Corollary. *If U and V are finitely generated subgroups of the free group F then $U \cap V$ is also finitely generated.* □

Given where this course is being held, it is worth formulating now the

Hanna Neumann Conjecture. *If u and v and t are the ranks of U, V and $U \cap V$ above and $u, v \geq 2$, then $t - 1 \leq (u-1)(v-1)$.*

With an extra factor 2 on the right this inequality was proven by Hanna Neumann about 40 years ago. Getting rid of the factor of 2 has proved remarkably difficult, although some partial results are known.

Rationality of subgroups seems to be the "right" property in many ways. When studying fundamental groups of hyperbolic manifolds, there is a fundamentally important geometric property called "geometric finiteness." Roughly, it refers to the existence of a finite sided polyhedral fundamental domain. By Milnor's Theorem (§1.2) the fundamental group of a compact hyperbolic manifold is a hyperbolic group. Swarup showed that a subgroup of such a group is a rational subgroup if and only if it is geometrically finite. In his 1995 thesis Lawrence Reeves has proved the analogous result if the manifold is only of finite volume rather than compact: in this case its fundamental group is not hyperbolic, but a subgroup is geometrically finite if and only if it is rational for some biautomatic structure on the group.

If a compact hyperbolic 3-manifold group $G = \pi_1 M$ has a subgroup isomorphic to a surface group, then that surface group is non-rational if and only if, after replacing our groups by subgroups of finite index if necessary (i.e., going to a finite cover of M), that surface group is the fundamental group of the fiber of a fibration of M. Thurston has conjectured that such fibered covers of M always exist. Equivalently, every hyperbolic 3-manifold group has non-rational surface subgroups.

7.3 Subgroups of bi-automatic groups

There is a beautiful theory of rational and other subgroups in bi-automatic groups due to Gersten and Short **[GS]**. We have already proved the main tool of this theory in §5.7: we saw that if g, g' are elements of a biautomatic group G then $\{h : h^{-1}gh = g'\}$ is a rational subset of G. Taking $g = g'$ this tells us that the centralizer subgroup $Z_G(g) := \{h \in G : gh = hg\}$ of an element of G is always a rational subgroup. Now the centralizer of a finitely generated subgroup of G is the intersection of the centralizers of a set of generators of the subgroup, so we have proved:

Theorem. *If H is a finitely generated subgroup of a biautomatic group G then the centralizer $Z_G(H)$ is rational (and hence biautomatic).* □

Applying this to $H = G$ we see that the centre of a biautomatic group is a rational subgroup. Since the centre of the centralizer of an abelian subgroup H is an abelian group containing H, it follows that:

Lemma. *Any finitely generated abelian subgroup of a biautomatic group is contained in a rational abelian subgroup.* □

Now let G be a group with finite generating set $\mathcal{G}$. We say an element $g \in G$ has ***positive translation length*** if $\liminf_{n\to\infty} \ell(g^n)/n > 0$. We can also speak of positive translation length with respect to an automatic structure L, using word-length of L-words in place of word metric. However, since L-words are quasi-geodesic, this gives the same concept of positive translation length. In particular, it follows that an element of a rational subgroup H of an automatic group G has positive translation length in H if and only if it does so in G. The above lemma now easily implies:

Theorem. *Any element of infinite order in a biautomatic group has positive translation length.* □

This yields strong restrictions on what subgroups biautomatic groups (and hence also hyperbolic groups) can have. For example, most Baumslag-Solitar groups are ruled out as subgroups and also:

Corollary. *A nilpotent or polycyclic subgroup of a biautomatic group must be virtually abelian.*

The proof is basicly that if not, one could find an element of infinite order with zero translation length in the subgroup, and hence certainly with zero translation length in the group.

8 Almost Convexity

Almost convexity is a geometric property due to Cannon **[C2]** which gives a highly efficient way to build the Cayley graph.

As usual, we take $\Gamma = \Gamma_{\mathcal{G}}(G)$ to be the Cayley graph of G, and let $B(n) = \{x \in \Gamma \mid d(1,x) \le n\}$ and $S(n) = \{x \in \Gamma \mid d(1,x) = n\}$. We are most specifically thinking of Γ as a bona fide metric space here.

We say Γ is ***almost convex*** if there is a constant K so that if $g, g' \in B(n)$ with $d(g,g') \le 2$ then there is a path p inside $B(n)$ running from g to g' with $\mathrm{len}(p) \le K$.

We say Γ is ***almost convex*** (k) if there is a constant $K(k)$ so that if $g, g' \in B(n)$ with $d(g,g') \le k$ then there is a path p inside $B(n)$ running from g to g' with $\mathrm{len}(p) \le K(k)$.

Thus Γ is almost convex if it is almost convex (2). The reason for this is the following theorem which is a straight forward exercise in the definition.

Theorem. *If* Γ *is almost convex* (2) *then* Γ *is almost convex* (k) *for all* k.

Let us look at some consequences.

Theorem. *Suppose* Γ *is almost convex. Then* G *is finitely presented and has at worst exponential isoperimetric inequality.*

In fact, we shall see that G can be presented as $\langle \mathcal{G} \mid \mathcal{R} \rangle$ where

$$\mathcal{R} = \{w \in \mathcal{G}^* \mid \overline{w} = 1 \text{ and } \operatorname{len}(w) \leq K + 2\}$$

where K is the almost convexity constant of Γ.

Proof. Suppose w is a word evaluating to 1 in G. We consider w as a closed path based at 1, and let n be minimal so that w lies in $B(n)$. Write $w = w_0 = u_1 v_1 \ldots u_k v_k$ where the u_i are the portions of w escaping $B(n-1)$. (So unless $n = 1$, u_1 is empty!) Now observe that each u_i consists of either a single edge whose midpoint is at distance $n - \frac{1}{2}$ from the identity, or a pair of edges whose common vertex is at distance n from the identity. For if two consecutive vertices of u_i are outside $B(n-1)$, then the midpoint of their edge is outside $B(n)$. Hence we can replace each u_i with a path p_i giving $w = w_1 = p_1 v_1 \ldots p_k v_k$ so that

- each p_i has length at most K
- $\overline{w_1} = \overline{w_0}$ and
- w_1 lies entirely inside $B(n-1)$.

Continuing in this way produces w_n lying entirely inside $B(0)$, that is, w_n is the empty word.

Now each step in this process corresponds to the application of relators of the form $u_i^{-1} p_i$ and each of these has length at most $K + 2$. You must now check that this corresponds to the creation of a Van Kampen diagram for w. Furthermore, in going from w_i to w_{i+1} we used at most $\operatorname{len}(w_i)$ relators and increased length by at most a factor of K. The exponential isoperimetric inequality now follows. $\square$

There is a highly efficient algorithm for building an almost convex Cayley graph.

Theorem. *If Γ is almost convex then there is an algorithm which produces the edges and vertices of Γ at a constant rate.*

In some sense, this theorem is only morally true. If you wish to carry out such an algorithm in a hard core Turing machine fashion, you must keep lists of edges and vertices. As the lengths of these lists grow and the lengths of the names of the things on these lists grow, you will spend worse than linear time simple traversing your list. We should not concentrate here on these hard core machine-theoretic impedimentia.

Proof. We seek to build $B(n)$. We can certainly do this for $n = 0$! Suppose now that we have built $B(n-1)$, and that for convenience, we have kept track of $S(n-1)$. Now each vertex of Γ has emanating from it one edge from each element of $\mathcal{G}$. For each element of $B(n-2)$ we have already found all its edges. For each $g \in S(n-1)$ let E_g be the set of edges of g not in $B(n-1)$. We form a "proposed $B(n)$" by appending to $B(n-1)$ the missing edges $E(g)$ for each $g \in S(n-1)$. Call this object $X(n)$. Now it is clear that $B(n)$ is a quotient of $X(n)$. That is, to find $B(n)$, we must determine for each pair of edges e and e' in $X(n) - S(n-1)$ whether

- e and e' are the inverse edges (and thus both have endpoints already found in $S(n-1)$ or
- e and e' have a common endpoint in $S(n)$. After determining this and making appropriate identifications on $X(n)$ we have found $B(n)$.

So suppose e and e' emanate respectively from g and g'. Then almost convexity ensures that if either of the two above situations occur then there is a short path p from g to g' in $B(n-1)$ (which we have already found) which suffices to determine this fact. Thus verifying each new vertex of $B(n)$ requires only a fixed amount of checking in $B(n-1)$.

It turns out that almost convexity is not a group property.

Theorem (Thiel) [T]. *Let H_{2n+1} be the Heisenberg group of dimension $2n+1$:*

$$\langle a_1, \ldots, a_n, b_1, \ldots, b_n, c : [a_i, b_i] = c, [a_i, b_j] = 1 \text{ for } i \neq j, c \text{ central} \rangle.$$

Then for $n > 1$ there are generating sets $\mathcal{G}_{2n+1}$ and $\mathcal{G}'_{2n+1}$ for H_{2n+1} so that $\Gamma_{\mathcal{G}_{2n+1}}(H_{2n+1})$ is almost convex, but $\Gamma_{\mathcal{G}'_{2n+1}}(H_{2n+1})$ is not.

Those groups that are almost convex include
- hyperbolic groups (in any generating set),
- virtually abelian groups (in any generating set),
- groups with geodesic automatic structure,
- $\pi_1(M)$ where M is a closed 3-manifold with a uniform geometric structure which is not solvgeometry Finitely presented groups which are not almost convex in any generating set include fundamental groups of closed solvegeometry 3-manifolds **[CFGT]** and solvable Baumslag-Solitar groups **[MS]**.
  There is much to be learned about this property.

9 Growth functions and growth rates

Let G be a group with finite generating set $\mathcal{G}$. We denote by $S(n)$ the set of group elements at distance n from the identity and let $s(n) = |S(n)|$. The ***growth function*** of G with respect to $\mathcal{G}$ is the power series

$$f(t) = \sum_{n=0}^{\infty} s(n) t^n = \sum_{g \in G} t^{\ell(g)},$$

where $\ell(g) = d(g, 1)$. If this is the power series of a rational function, one says that G has ***rational growth*** with respect to $\mathcal{G}$. In any case this function has a positive radius of convergence since $s(n) < |\mathcal{G}|^n$.

Theorem. *If G has a geodesic automatic structure with respect to $\mathcal{G}$ then its growth with respect to $\mathcal{G}$ is rational.*

Proof. We can find a bijective geodesic automatic structure, and then the result is immediate from the following standard result. □

Let $L \subset \mathcal{A}^*$ be a language. Its growth function is the power series whose coefficients are the number of words of length n in the language.

Theorem. *The growth of a regular language L is rational.*

Proof. If $\mathcal{A}$ is a finite state automaton for L one forms the transition matrix M for $\mathcal{A}$ whose rows and columns are indexed by the states of $\mathcal{A}$ and whose entry m_{ij} counts the number of edges from state i to state j. This matrix is the adjacency matrix for the underlying graph of $\mathcal{A}$. Notice that the (i,j)-entry of M^n counts the number of paths of length n from i to j. Thus the number of words of length n in L is $v_1 M^n v_2$ where v_1 is the row vector with a 1 at the start state and 0's elsewhere and v_2 is the column vector with 1's at accept states and 0's elsewhere. The growth function is therefore given by the function $v_1\big(\sum_{i=0}^{\infty}(tM)^i\big)v_2$. This function is rational because it can be rewritten as $v_1(I-tM)^{-1}v_2$ and $(I-tM)^{-1}$ will be a matrix whose entries are rational functions of t (see, e.g., **[C1]**). □

By a modification of this argument one can also show:

Theorem. *If $\Gamma_{\mathcal{G}}(G)$ has finitely many cone types (in particular, if it has the falsification by fellow traveler property) then G has rational growth with respect to $\mathcal{G}$.*

It is unknown if language of geodesics being regular implies rational growth, although the above result is very close to this. There exist groups with rational growth with respect to a given generating set for which the language of geodesics is not regular. For example, virtually abelian groups have rational growth with respect to any finite generating set but need not have regular language of geodesics.

It is an important open problem whether having rational growth is independent of finite generating set.

9.2 Growth rate

Another important invariant of a finitely generated group is the growth rate of the function $s(n)$, e.g., the question of whether it is polynomial or not, and the degree of polynomial growth if it is polynomial. Its growth cannot be more than exponential, in contrast to the isoperimetric function. There do however exist groups with growth strictly between polynomial and exponential; whether they can be finitely presented is still open.

The most important theorem in the area is:

Theorem (Gromov). *A finitely generated group has polynomial growth if and only if it is virtually nilpotent.*

A group is nilpotent if there is a k such that G_k is trivial, where $G_0 = G$ and $G_i = G_{i-1}/Z(G_{i-1})$. The proof uses a beautiful construction of “asymptotic cone" of a group which has had other applications, but which we cannot go into here.

References

The following are neither exhaustive nor intended to establish historic credit.

[ABC] J.M. Alonso, T. Brady, D. Cooper, V. Ferlini, M. Lustig, M. Mihalik, M. Shapiro, H. Short, H. Short, ed., Notes on word hyperbolic groups, in Group Theory from a Geometric Viewpoint, E. Ghys, A. Haefliger, A. Verjovsky eds., World Scientific, (1991) 2-63.

[BBMS] G. Baumslag, M. Bridson, C.F. Miller, H. Short, Finitely presented subgroups of automatic groups and their isoperimetric functions, Bull. L.M.S. to appear.

[BGSS] G. Baumslag, S. M. Gersten, M. Shapiro and H. Short, Automatic groups and amalgams, Journal of Pure and Applied Algebra **76** (1991), 229–316.

[B] B. Bowditch, Notes on Gromov's hyperbolicity criterion for path metric spaces. in Group Theory from a Geometric Viewpoint, E. Ghys, A. Haefliger, A. Verjovsky eds., World Scientific, (1991) 64–167.

[BH] B. Brink, R. Howlett, Math. Annalen (about 1993/4).

[C1] J. Cannon, The combinatorial structure of cocompact discrete hyperbolic groups, Geom. Dedicata **16** (1984), 123–148.

[C2] J. Cannon, Almost convex groups, Geometrae Dedicata **22**, 197—210 (1987).

[C3] J. Cannon, The theory of negatively curved spaces and groups in Ergodic Theory, Symbolic Dynamics and Hyperbolic Spaces, T. Bedford, M. Keane, C. Series (eds.) Oxford University Press (1991) 315-369.

[CFGT] J. Cannon, W. Floyd, M. Grayson, W. Thurston, SolvGeometry groups are not almost convex, Geom. Dedicata ca.1989.

[CS] D. Cartwright, M. Shapiro, Hyperbolic buildings, affine buildings, and automatic groups, Michigan Math. J. 42 (3) (1995 or 6).

[Ch] R. Charney, Artin groups of finite type are biautomatic, Mathematische Annalen 292 (1992), 671–683 (see also: R. Charney, Geodesic automation and growth functions for Artin groups of finite type, to appear in Mathematische Annalen.)

[CDP] M. Coonaert, T. Delzant, A. Papadopoulos, Géometrie et théorie des groupes, Les groupes hyperboliques de Gromov, Lecture notes in Math. 1441 Springer Verlag 1990.

[DS] M. Davis and M. Shapiro, Coxeter groups are almost convex, Geometriae Dedicata **39** (1991) 55–57.

[D] M. Dehn, Papers on Group Theory and Topology, translated and introduced by J Stilwell, Springer-Verlag 1987

[Du] M. Dunwoody, The accessibility of finitely presented groups, Inv. Math. (1985), 449–457

[ECHLPT] D.B.A. Epstein, J.W. Cannon, D.F. Holt, S.V.F. Levy M.S. Paterson, and W.P. Thurston, “Word Processing in Groups,” Jones and Bartlett Publishers, Boston, 1992

[G] S. Gersten, Dehn functions and l_1 norms of finite presentation, in "Algorithms and Classification in Combinatorial Group Theory". MSRI Publications No.23, edited by Baumslag and C.F. Miller, Springer Verlag, 1991.

[GS1] S.M. Gersten and H. Short, Rational sub-groups of bi-automatic groups, Annals of Math. **134** (1991), 125–158.

[GS2] S.M. Gersten and H. Short, Small cancellation theory and automatic groups I, II, Invent. Math. 102 (1990), 305–334, 105 (1991), 641–662.

[GH] E. Ghys and P. de la Harpe, Sur les Groupes Hyperboliques d'après Mikhael Gromov, Birkauser Progress in Mathematics, vol. 83 (1990)

[Gr1] M. Gromov, Hyperbolic Groups, Essays in Group Theory, MSRI series vol. 8, S. M. Gersten, ed., Springer-Verlag

[Gr2] M. Gromov, Asymptotic invariants of infinite groups, Geometric Group Theory Vol2, Ed: G. Niblo and M. Roller, Cambridge Univ. Press (1993)

[L] I. G. Lysenok, On some algorithmic properties fo hyperbolic groups, Izv. Akad. Nauk. SSSR ser. Math 53 (4) (1989); English transl. in Math. USSR Izv. 35 (1990), 145-163.

[LS] R. Lyndon and P. Schupp, *Combinatorial Group Theory*, Springer Verlag 1977.

[MS] C.F. Miller, M. Shapiro, unpublished.

[M] J. Milnor, A note on curvature and fundamental group, J. Diff. Geom. **2** (1968), 1–7.

[Mo1] L. Mosher, Mapping class groups are automatic, Annals of Math (to appear).

[Mo2] L. Mosher, Central quotients of biautomatic groups, Preprint.

[NR] W.D. Neumann and L. Reeves, Central extensions of word hyperbolic groups, Annals of Math. (to appear 1996).

[NS1] W.D. Neumann and M. Shapiro, Equivalent automatic structures and their boundaries, Internat. J. Alg. Comp. **2** (1992), 443–469

[NS2] W.D. Neumann and M. Shapiro, Automatic structures, rational growth, and geometrically finite hyperbolic groups, Invent. Math. **120** (1995), 259–287.

[Pa1] P. Papasoglu, On the subquadratic isoperimetric inequality, in "Geometric Group Theory" Charney/Davis/Shapiro Eds. Ohio State University Math. Research Inst. Publications 3 (de Gruyter 1995), 149–158.

[Pa2] P. Papasoglu, Strongly Geodesically Automatic Groups Are Hyperbolic, Invent. Math 12 (1995), 323–334.

[Se] Z. Sela Structure and rigidity in (Gromov) Hyperbolic groups ..., Annals of Math (1995).

[S1] M. Shapiro, Automatic structures and graphs of groups, in "Topology '90" (Walter de Gruyter Verlag 1992), 355–380.

[S2] M. Shapiro, unpublished, announced in **[S1]**

[T] C. Thiel, Zur Fast-Konvexität einiger nilpotenter Gruppen, Bonner Mathematische Schriften, 1992.