Diagrams and groups

Hamish Short

June, 2006

0.1 Introduction

These notes are intended to be an introduction to the use of diagrams to study problems in the domain of finitely presented groups, in particular decision problems such as the word problem. The main topics are: van Kampen diagrams, isoperimetric inequalities, small cancellation theory and hyperbolic groups. To illustrate the ideas some examples will be studied in detail: a polynomial lower bound for the isoperimetric inequality for free nilpotent groups (studied by Tim Riley in his course), and a polynomial upper bound for the isoperimetric inequality for certain normal subgroups of hyperbolic groups. Details concering some of the topics touched on during the course are given here : I would like to thank Tim Riley, Noel Brady and all the others who have helped me to improve the preliminary version of these notes. Many thanks are due to the Centre de Recerca Matemàtica, and in particular to the organisers José Burillo and Enric Ventura, for their invitation to give this course, and for ensuring the smooth running of the event. And of course one is always grateful to the members of the audience for making the workshop a lot of fun.

In the first chapter, we shall see how the Cayley graph of a finitely generated group gives a geometric object providing a language in which to talk about many of the properties of the group. Geometric group theory studies properties of the Cayley graphs of groups.

In the second chapter we describe van Kampen's diagrams which provide a method for visualising relations in presentations of groups. We show how to obtain such diagrams and their dual pictures. We shall give some generalisations and applications to free products and HNN extensions.

Small cancellation theory gives a method of working with certain restricted forms of finite presentations; this is studied in chapter 3. When a finite presentation satisfies certain easily verifiable conditions, the word problem is solvable in a particularly simple way. This theory has its origins in Dehn's original work, and led to Gromov's definition of word hyperbolic groups.

In chapter 4 we give some details about quasi–isometries and show that a quasi–isometry of Cayley graphs preserves the property of being finitely presented, the property of having a solvable word problem, and the type of isoperimetric inequality satisfied. We shall describe some properties of word hyperbolic groups.

A certain method for obtaining lower bounds for isoperimetric inequalities is described in chapter 5. This has an application to nilpotent groups, where this polynomial (of degree c=nilpotency class) bound can be combined with the ($c+1$-degree) polynomial upper bound (see chapter 4 of Tim Riley's notes) in the case of free nilpotent groups.

Finally we show how to obtain a polynomial isoperimetric inequality for certain normal subgroups of hyperbolic groups which are cyclic extensions. This applies to certain examples of Noel Brady.

Unfortunately time did not permit the covering of other topics, in particular Weinbaum's proof [30] of the conjugacy problem for alternating knots and links, Gromov's version of small cancellation theory (see Ollivier's paper [26]), and Klyachko's [19] work on the Kervaire conjecture (see also [12]).

Chapter 1

Dehn's problems and Cayley graphs

We shall suppose known the basic definition and properties of a free group — some of the many references available for this, in particular some favouring a geometric approach, are the books by Magnus, Karrass and Solitar [21], Lyndon and Schupp [LS], Bridson and Haefliger [6], Ghys and de la Harpe [14] and Hatcher [17].

We shall use $F(\mathcal{A})$ to denote the free group on $\mathcal{A}$. Let $R \subset F(\mathcal{A})$, and let $\langle\langle R\rangle\rangle$ denote the subgroup normally generated by R, i.e. the intersection of all normal subgroups which contain R. This is of course a normal subgroup, and it is not hard to see that it can be described as:

$$\langle\langle R\rangle\rangle = \{\prod_{i=1}^{M} p_i {r_i}^{\epsilon_i} {p_i}^{-1} \mid \forall M \in \mathbb{N}, \forall p_i \in F(\mathcal{A}), \forall r_i \in R, \forall \epsilon_i = \pm 1\}$$

Let Γ be a group, and $\mathcal{A}$ a generating set for Γ. In the usual naïve sense, this means that $\mathcal{A}$ is a subset of Γ (In this sense, the trivial group could only have one element in a generting set.) Here this will mean that there is a surjective homomorphism $\Phi : F(\mathcal{A}) \twoheadrightarrow \Gamma$. (Thus the trivial group can have a large generating set.) Any word w in (i.e. finite product of) the generators and their inverses thus represents an element of Γ; the length of w we write $\ell(w)$, meaning the number of generators and their inverses appearing in the product. The obvious shortenings of this product, meaning the removal of subwords of the form aa^{-1} and $a^{-1}a$ for $a \in \mathcal{A}$ are called *reductions*, and the word is reduced if none are possible.

The kernel $\ker\Phi$ is a normal subgroup; if $R \subset F(\mathcal{A})$ is a subset which normally generates $\ker\Phi$, i.e. $\langle\langle R\rangle\rangle = \ker\Phi$, then say that R is a set of relators for Γ with respect to the generating set $\mathcal{A}$. Such a set R always exists — it suffices to take $R = \ker\Phi$. What is more interesting is to try to obtain, if possible, a finite set R, or if not, some recursive or "systematic" set R. The corresponding presentation of Γ is written $\langle \mathcal{A} \mid R\rangle$ or $\langle \mathcal{A}; R\rangle$, meaning that the map $\mathcal{A} \to \Gamma$

induces an isomorphism $F(\mathcal{A})/\langle\langle R \rangle\rangle \to \Gamma$. The group Γ is finitely generated if it has a presentation with $\mathcal{A}$ finite, and is finitely presentable (or finitely presented) if it has a presentation with both $\mathcal{A}$ and R finite. We shall always assume that the words in R are cyclically reduced, as their cyclic reduction does not change the normal subgroup that is generated.

In 1912, Max Dehn [10] (this is available in an English translation, thanks to Stillwell) posed the three algorithmic problems for finitely presentable groups at the base of combinatorial group theory. It is worth noting that he did this well before Turing and Gödel's work, though in the spirit of Hilbert's problems. It was not until the 1950's that it was proved that in general such algorithms do not exist (by Novikov and Boone). Here are the three problems in their original formulation (Stillwell's translation [10, pages 133-134]) (see Figure 1.1 for the original):

The Word Problem: An element of the group is given as a product of generators. One is required to give a method whereby it may be decided in a finite number of steps whether this element is the identity or not.

The Conjugacy Problem: Any two elements S and T of the group are given. A method is sought for deciding the question whether S and T can be transformed into each other, i.e. whether there is an element U of this group satisfying the relation $S = UTU^{-1}$.

The Isomorphism Problem: Given two groups, one is to decide whether they are isomorphic or not (and further whether a given correspondence between the generators of one group and elements of the other group is an isomorphism or not).

1. *Das Identitätsproblem:* Irgend ein Element der Gruppe ist durch seine Zusammensetzung aus den Erzeugenden gegeben. Man soll eine Methode angeben, um mit einer endlichen Anzahl von Schritten zu entscheiden, ob dies Element der Identität gleich ist oder nicht.

2. *Das Transformationsproblem:* Irgend zwei Elemente S und T der Gruppe sind gegeben. Gesucht wird eine Methode zur Entscheidung der Frage, ob S und T ineinander transformiert werden können, d. h. ob es ein Element U der Gruppe gibt, welches die Relation befriedigt:

$$S = UTU^{-1}.$$

3. *Das Isomorphieproblem:* Zwei Gruppen sind gegeben, man soll entscheiden, ob sie isomorph sind oder nicht (und, des weiteren, ob eine gegebene Zuordnung der Erzeugenden der einen Gruppe zu Elementen der andern Gruppe eine isomorphe Zuordnung ist oder nicht).

Figure 1.1: Dehn's three decision problems

An extremely efficient solution exists for certain finite presentations which have a *Dehn Algorithm*. We say that a presentation has such an algorithm when any word $w \in \langle\langle R \rangle\rangle$ always contains more than half of some relator (considered cyclically): i.e. w is a word in the generators of the form $Ur'V$ and there is some (cyclic conjugate of some) $r \in R \cup R^{-1}$ such that $r = r'r''$ and $\ell(r'') < \ell(r')$. If this is the case, then the group element represented by the subword r' is equal in the group to the element represented by r''^{-1}, and replacing r' by this shorter word reduces the length of w. Continuing in this way, w is trivial if and only if this procedure of looking for one of the finite number of long subwords of the cyclic conjugates of the relators, and replacing it by shorter word to give an element equal to w in the group, eventually leads to the empty word. If a finite presentation has this property then the group presented is word hyperbolic, and any word hyperbolic group has such a finite presentation [2] (this was originally pointed out by Jim Cannon — in fact it has such a presentation with respect to any finite generating set: it suffices to add enough relators, see for instance [2]).

The first two problems describe a property of a finitely presentable group: if there is such an algorithm for the finite presentation $\mathcal{P}_1$ of Γ, and $\mathcal{P}_2$ is another finite presentation of Γ, then there is algorithm to solve the problem over $\mathcal{P}_2$ (as one can find an isomorphism between the presentations when one knows that one exists, by simply enumerating all Tietze transformations).

The third problem is so badly unsolvable in general that it is impossible to give an algorithm to recognize presentations of the trivial group. This is despite the fact that there is a procedure to enumerate all presentations of the trivial group (via Tietze transformations).

There is an obvious enumeration of $\langle\langle R \rangle\rangle$ by using the usual diagonal method on the lists of different numbers of conjugates of elements of $R^{\pm 1}$, ordered by the list of conjugating elements $p_i \in F(\mathcal{A})$. The hard part of the word problem resides in detecting words which represent <u>non–trivial</u> elements of Γ.

Given the enumeration procedure described above, if we know that the word w does indeed represent the trivial element in the presentation, the expression for w as a product of conjugates of relators $w = \prod_{i=1}^{M} p_i r^{\epsilon_i} {p_i}^{-1}$ can be found, where $\epsilon_i = \pm 1$ and $M \in \mathbb{N}$. The smallest such number M is called the <u>area</u> of w.
The function $\delta_{\mathcal{P}} : \mathbb{N} \to \mathbb{N} : \delta_{\mathcal{P}}(n) = \max_{\{w \in \langle\langle R \rangle\rangle, \ell(w) \leq n\}} Area_{\mathcal{P}}(w)$ is called the Dehn function of the presentation $\mathcal{P}$. An *isoperimetric inequality* for the presentation is a function $f : \mathbb{N} \to \mathbb{R}$ such that for all $n \in \mathbb{N}$, $\delta_{\mathcal{P}}(n) \leq f(n)$. We shall study the dependence of these functions on the actual presentation later.

Theorem 1.1. *A finite presentation satisfies a recursive isoperimetric inequality if and only if it has a solvable word problem.*

Proof. If the word problem is solvable, then for each $n \in \mathbb{N}$, and each word w of length n, it is possible to decide whether or not w lies in $\langle\langle R \rangle\rangle$. If it does, then the enumeration procedure above eventually gives some expression for w as a product of conjugates of relators. In this way, examining all words of length at most n, this gives an upper bound for the Dehn function $\delta_{\mathcal{P}}(n)$ as required.

If a recursive function f bounding the Dehn function is know, and $w \in F(\mathcal{A})$ of length n is given, then calculate $f(n)$. It remains to calculate all products of at most $f(n)$ conjugates of the relators and their inverses. *A priori* the lengths of the conjugating elements which need to be tried is not bounded, and so all elements of $F(\mathcal{A})$ must be tried. We shall see later 2.2, using van Kampen diagrams, that it suffices to check conjugating elements of length at most $f(n) \max_{r \in R} \ell(r) + \ell(w)$, so that there is a finite number of combinations which must be checked. This is the only step which is not immediate and it is best seen from the diagrams which are to be introduced in chapter 2. □

In the 1980's Gromov introduced a class of groups generalising discrete groups of isometries acting cocompactly on hyperbolic spaces. It is an interesting exercise to read Gromov's papers about hyperbolic groups alongside Dehn's articles about the conjugacy problem for surface groups (for instance in Stillwell's translation, see [10]). There are several equivalent definitions of the class of hyperbolic groups, some of which we shall explore later. One definition is in terms of area functions:

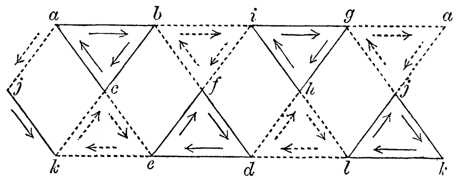
Definition 1.2. A finitely presentable group Γ is *word hyperbolic* if it has a finite presentation which satisfies a linear isoperimetric inequality.

In fact a group is word hyperbolic if and only if it satisfies a sub–quadratic isoperimetric inequality [4],[27]. Also, as we noted earlier, a group is word hyperbolic if and only if it has a finite presentation which has a Dehn algorithm (see for instance [2]). It is easy to solve the word and conjugacy problems for finitely generated abelian groups, or free groups with respect to free generating sets. The first case gives a quadratic isoperimetric inequality, the second a linear one.

The idea of representing a group by a graph goes back to Cayley, though he only uses them in the context of finite groups. Dehn extends the ideas to infinite groups and uses them (*Gruppenbilder*) intensively to study fundamental groups of closed compact surfaces.

Definition 1.3 (Cayley graph)**.** Let $\mathcal{A}$ be a generating set for the group Γ. The Cayley graph of Γ with respect to $\mathcal{A}$, written $Cay^1(\Gamma, \mathcal{A})$, has a vertex for each element $g \in \Gamma$, and for each such vertex g_1, and each $a \in \mathcal{A}$, there is an oriented labelled edge from the vertex g_1 to the vertex g_2 if and only if $g_2 = g_1.a$ in Γ (which we write $g_2 =_\Gamma g_1 a$). Notice that it may be that $g_1 =_\Gamma g_2$ when the generator a represents the identity element of Γ. The fact that $\mathcal{A}$ is a generating set means that this graph is connected. Assign length 1 to each edge to consider $Cay^1(\Gamma, \mathcal{A})$ as a metric space, where the distance $d_{\mathcal{A}}(v, v')$ between the points v, v' is the length of the shortest path between them.

On Γ this defines an integer–valued metric on Γ called the word–metric (it depends on the choice of generating set). As is usual in the context of labelled oriented graphs, to a path between vertices in the Cayley graph is associated a word in the free group $F(\mathcal{A})$ (this word may be unreduced). The word is obtained by writing the letter corresponding to the label on each edge in the order traversed, I represent this by a diagram, the lines of which were red and black, and they



will be thus spoken of, but the black lines are in the woodcut continuous lines, and the red lines broken lines: each face indicates a cyclical substitution, as shown by the arrows. The figure should be in the first instance drawn with the arrows, but without the letters, and these may then be affixed to the several points in a perfectly arbitrary manner; but I have in fact affixed them in such wise that the group given

Figure 1.2: From Cayley's paper [7]: a Cayley graph the alternating group A_4, with presentation $\langle x, y \mid x^3, y^3, (xy)^3\rangle$ (the vertices a, j, k are repeated).

with the exponent ± 1 according to whether the direction of the path agrees with (+1) or is opposite to (-1) the orientation of the edge.

Some elementary properties of Cayley graphs :

• The group Γ acts freely on the left on $Cay^1(\Gamma, \mathcal{A})$ by isometries with respect to the distance $d_{\mathcal{A}}$, and transitively on the set of vertices.

• The Cayley graph $Cay^1(\Gamma, \mathcal{A})$ of a finitely generated group is a covering space of the 1–complex $K(\mathcal{A})$, with one vertex and $|\mathcal{A}|$ edges (each forming a loop), whose fundamental group is $F(\mathcal{A})$. We can regard $K(\mathcal{A})$ as $Cay^1(1, \mathcal{A})$, a Cayley graph of 1, the trivial group.

• The usual correspondence between covering spaces and subgroups of the fundamental groups, says that when $\langle \mathcal{A} \mid R\rangle$ is a presentation (finite or infinite) of the group Γ, the Cayley graph $Cay^1(\Gamma, \mathcal{A})$ is the cover of $K(\mathcal{A})$ corresponding to the normal subgroup $\langle\langle R\rangle\rangle$ of $F(\mathcal{A})$, the fundamental group of $K(\mathcal{A})$.

• In fact, if Γ' is a normal subgroup of Γ, then the quotient space $Cay^1(\Gamma, \mathcal{A})/\Gamma'$ is a Cayley graph of Γ/Γ'.

The diagram has the property that every route, leading from any one letter to itself, leads also from every other letter to itself; or say a route leading from a to a, leads also from b to b, from c to c,..., from l to l; and we can thus in the diagram speak absolutely (that is, without restriction as to the initial point) of a route as leading from a point to itself, or say as being equal to unity; *it is in virtue of this property* that the diagram gives a group.

Figure 1.3: Cayley's description of the group acting on its graph.

• Fixing some vertex v of $Cay^1(\Gamma, \mathcal{A})$ as a base point (for instance the vertex corresponding to the identity element 1), the word $w \in F(\mathcal{A})$ defines a unique path γ_w based at v. The word represents the identity element of Γ if and only if the path γ_w is a loop (note Cayley's remark on this, Fig. 1.3).

• To solve the word problem, it suffices to build the Cayley graph, or at least, to give an algorithm which, given a word of length n, constructs the ball of radius $n/2$ about the identity element in the Cayley graph, which is enough of it to see whether or not the word w labels a loop or not (this is basically the Todd–Coxeter algorithm).

Here are three essential, elementary, examples of cayley graphs. For more examples of Cayley graphs of two generator groups, seen as covering spaces, see [17, p. 58].

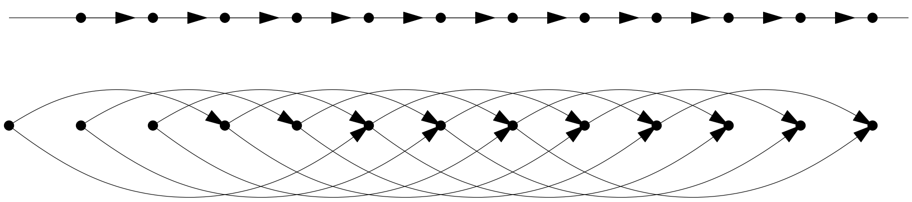


Figure 1.4: Two Cayley graphs for $\mathbb{Z}$ with generating sets $\{1\}$ and $\{3, 5\}$.

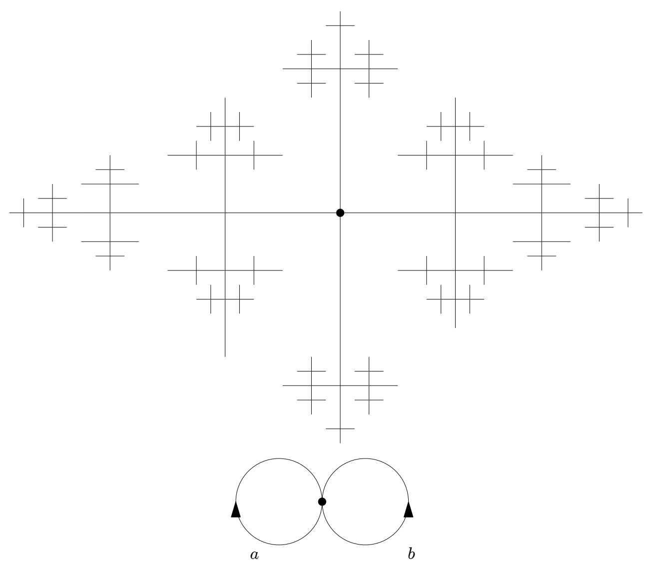


Figure 1.5: The Cayley graph of a free group on two generators with respect to a free basis; the a edges are horizontal, oriented from left to right, the b edges vertical oriented upwards.

Chapter 2

van Kampen Diagrams and Pictures

We now introduce the principal tool we shall use here for examining the word problem. These are diagrams introduced by Egbert van Kampen in 1933 [29].

Definition 2.1 (van Kampen or Dehn diagram)**.** Let $\mathcal{P} = \langle \mathcal{A} \mid R \rangle$ be a (usually finite) presentation for the group Γ. As is usual, we shall suppose that the relations in R are cyclically reduced. Let R^C denote the cyclic closure of R, which is the set of all cyclic conjugates of elements of R and their inverses:
$R^C = \{(p^{-1}rp)^{\pm 1} \mid p \text{ an initial segment of } r \in R\}$.

Let $\mathcal{D}$ be a finite, connected, oriented, based, labelled, planar graph where each oriented edge is labelled by an element of $\mathcal{A}$. The base point lies on the boundary of the unbounded region of $\mathbb{R}^2 - D$. Suppose in addition that for each bounded region (face) F of $\mathbb{R}^2 - \mathcal{D}$, the boundary ∂F (of the closure of F) is labelled by a word in R^C. This word is obtained by reading the labels on the edges as they are traversed, starting from some vertex on the boundary of F, in one of the two possible directions. Each label on the edge traversed is given a ± 1 exponent according to whether the direction of traversal coincides with, or is opposite to, the orientation of the edge. The choice of direction and starting point alters the word read by inversion and/or cyclic conjugation. The *boundary word* of the diagram D is the word w read on the boundary of the unbounded region of $\mathbb{R}^2 - \mathcal{D}$, starting from the base vertex. Then we say that $\mathcal{D}$ is a *van Kampen diagram* for the boundary word w over the presentation $\mathcal{P}$.

The diagram can also be viewed as a 2–complex, with a 2–cell attached to the graph (viewed as a 1–complex) for each bounded region. This constructs a combinatorial 2–complex, as we shall see below.

Usually we can suppose that w is a freely reduced word, though probably not cyclically reduced.

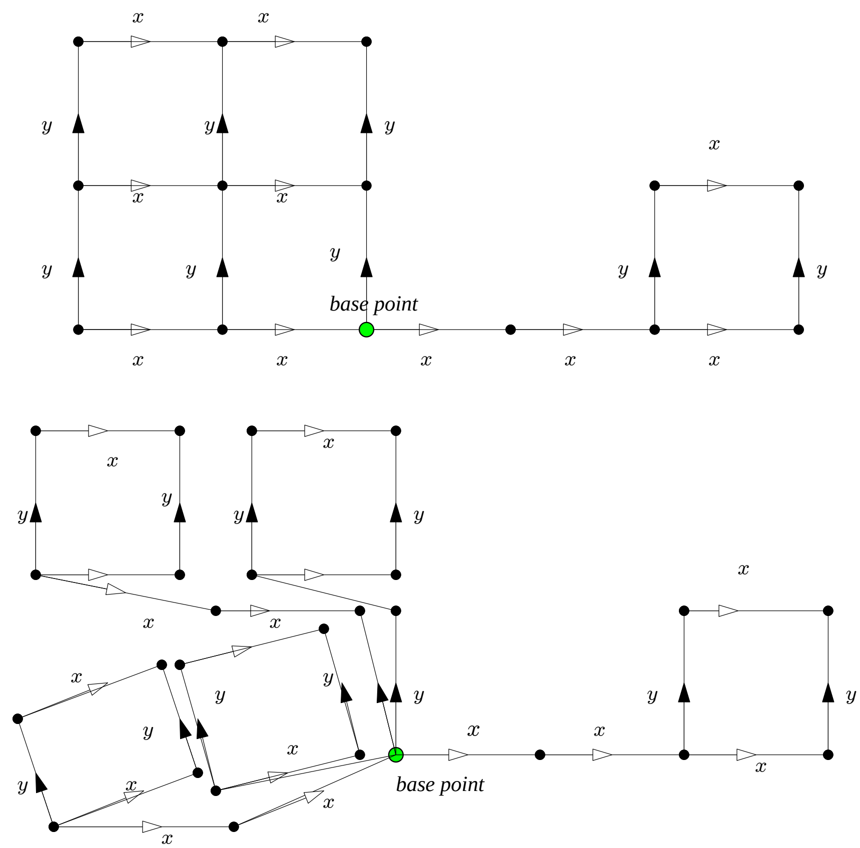


Figure 2.1: A diagram for $w = x^3yx^{-1}y^{-1}x^{-2}y^2x^{-2}y^{-2}x^2$ over the presentation $\langle x, y \mid xyx^{-1}y^{-1}\rangle$, and the same diagram for w deconstructed as $w = x^2rx^{-2}.yx^{-1}rxy^{-1}.yx^{-2}rx^2y^{-1}.x^{-1}rx.x^{-2}rx^2$, where $r = xyx^{-1}y^{-1}$

In 1933, Egbert van Kampen [29] defined his diagrammatic method of considering which words represent the identity element in the group given by a finite presentation. There he (essentially) stated the following result.

Theorem 2.2. *Let $\mathcal{P} = \langle \mathcal{A} \mid R\rangle$ be a presentation of the group $\Gamma = \Gamma(\mathcal{P})$.*
1) If $w \in \langle\langle R\rangle\rangle$, i.e. $w =_{\mathcal{P}} 1$, then there is a van Kampen diagram for w over $\mathcal{P}$.
2) If $\mathcal{D}$ is a van Kampen diagram for w over $\mathcal{P}$, then $w =_{\mathcal{P}}$.

We shall in establish something stronger, where we may allow bounded regions of the planar graph to have labels which are not freely (nor cyclically) reduced. This more general form will be useful when dealing with cancelling faces in 2.7.

Proposition 2.3. *Let $\mathcal{D}$ be a finite, connected, oriented, based, labelled, planar graph where each oriented edge is labelled by an element of $\mathcal{A}$. Suppose in addition that the bounded regions of $\mathbb{R}^2 - D$ are labelled by words whose freely reduced forms lie in $\langle\langle R\rangle\rangle$, and that the boundary label is w. Then a finite, connected, oriented, based, labelled, planar graph D' can be obtained from D such that all bounded regions are labelled by cyclically reduced words in $\langle\langle R\rangle\rangle$, and the boundary label is the reduced word corresponding to w.*

Proof. 1) We shall assume that w is given as a freely reduced word. Write $w = \prod_{i=1}^{M} p_i {r_i}^{\epsilon_i} {p_i}^{-1}$, with $r_i \in R$, $\epsilon_i = \pm 1$, $p_i \in F(\mathcal{A})$. This is an equality in the free group $F(\mathcal{A})$.

If $M = 1$ then draw in the plane a circle subdivided into $\ell(r_1)$ segments (or a regular polygon with $\ell(r_1)$ sides), and add an arc outside the circle at one of the vertices; subdivide the arc into $\ell(p_i)$ segments. Orient and label each of the edges appropriately. This is a van Kampen diagram for w (or rather for a word freely equal to w). We show below how to alter the diagram to obtain a diagram such that the boundary word is freely reduced.

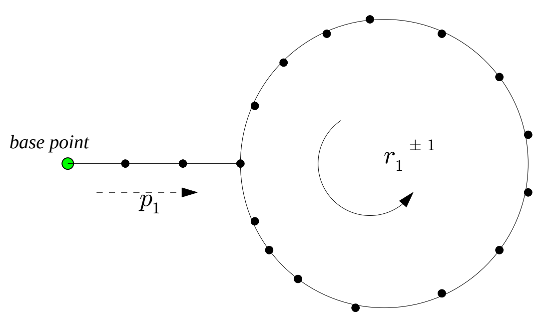


Figure 2.2: The case $M = 1$: w is the conjugate of a single relation

The general case: Draw M copies of the circle plus arc as before, all based at the same base point in the plane. Subdivide the arcs and the circles, and orient and label them appropriately.

When we assume that the relators in R are assumed to be cyclically reduced words, we can also ensure that the labels on all the boundaries of the regions are freely reduced words —a 1–dimensional reduction involving cancelling 1–cells.

If there is a vertex in the diagram other than the base vertex, which has valency one, then removing this vertex and the incident edge changes the label on the region having this edge on its boundary by a free reduction. If the valence one vertex is the base vertex, then removing this edge and vertex would correspond to a cyclic reduction of the boundary word w, which in general we do not allow, as we regard this word is being fixed (up to free reduction).

Suppose that there are two edges $e_1 = (v, v_1)$, $e_2 = (v, v_2)$ emanating from the same vertex $v \in \mathcal{D}$, both edges labelled by the same letter $x \in \mathcal{A}$, with the

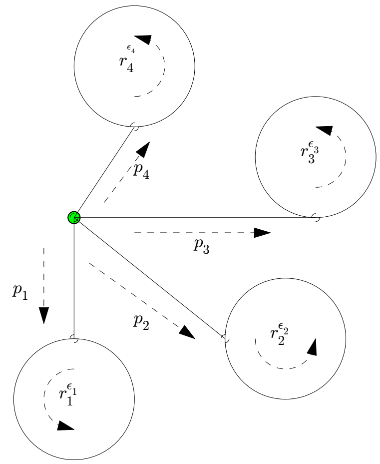


Figure 2.3: First step in the construction of a van Kampen diagram: general case.

same orientation with respect to v, and such that e_1 and e_2 are adjacent edges on the boundary of some face f of $\mathcal{D}$ (see figure 2.4). Identify the two edges e_1 and e_2, identifying the vertices v_1 and v_2. This changes the face f to the face f', with two fewer edges, and in the label on the boundary of this face, the cancelling letters $x^{-1}x$ are removed; all other face labels are unchanged.

Van Kampen says: *the two 1–cells can be brought into coincidence by a deformation without any other change in the complex*. There is however a problem, as there are four cases to consider: case 1: $v_1 \neq v \neq v_2$ and $v_1 \neq v_2$;
case 2: $v_1 = v \neq v_2$; case 3: $v_1 = v = v_2$; case 4: $v_1 = v_2 \neq v$.

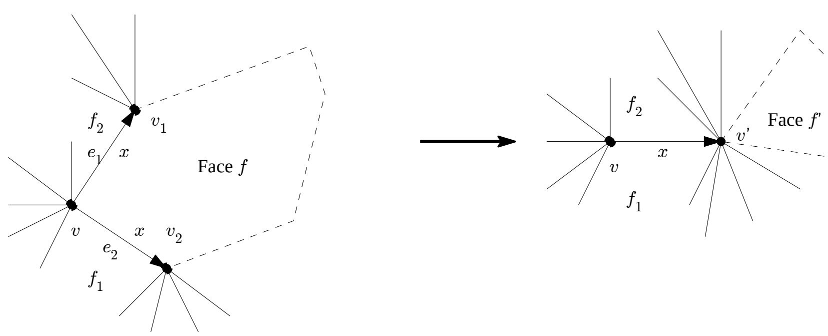


Figure 2.4: Case 1: edge identification when $v_1 \neq v \neq v_2$; here the face f is bounded. The identification can be realised by a map of the plane to the plane which essentially collapses a triangle onto an edge. Only one boundary label is affected, and that by a free reduction.

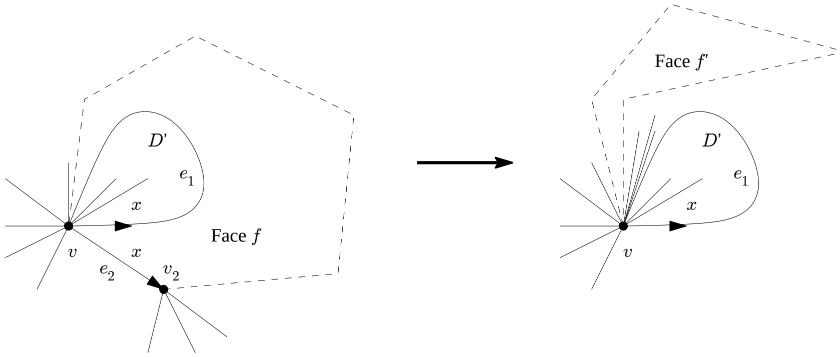


Figure 2.5: Case 2: $v_1 = v \neq v_2$. Again the identification can be realised by a map from the plane to the plane which collapses a triangle onto an edge, and only one boundary label is affected, as before.

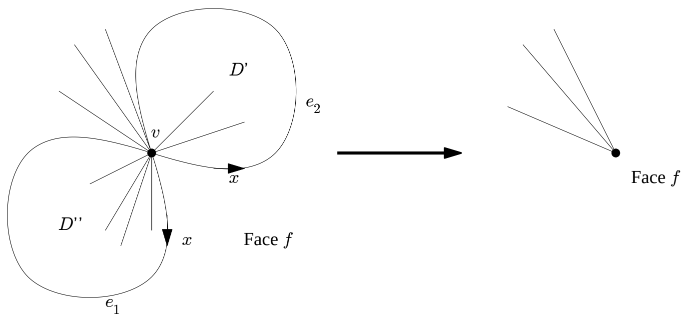


Figure 2.6: Case 3: $v_1 = v = v_2$. Here the identification of e_1 with e_2 cannot be realised while at the same time resting in the plane. However, we can realise the identification as a graph in the plane, together with a graph in a 2–sphere, made up of the two bounded regions D', D'' of the plane bounded by the edges e_1 and e_2. This 2–sphere is attached to the rest of the diagram at the vertex v. The 2–sphere is then discarded, leaving a diagram with fewer faces, and with the same labels on the remaining faces, except the face f where a free reduction, cancelling the adjacent letters x and x^{-1}, has been performed.

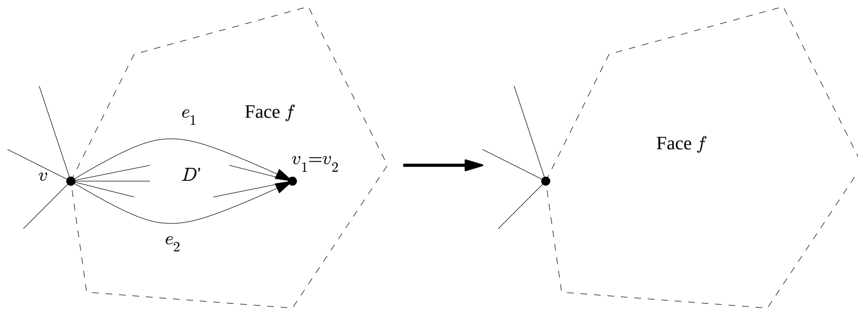


Figure 2.7: Case 4: $v_1 = v_2 \neq v$. Here the entire subdiagram D' enclosed by the two edges e_1 and e_2 is removed. Identifying the edges e_1 and e_2 produces a 2–sphere attached to the plane along the edge $e_1 = e_2$.

2) Suppose now that $\mathcal{D}$ is a van Kampen diagram, based at the vertex v, with M bounded faces. Let f be a bounded face which meets the unbounded region of the plane in a non-empty set B, containing an edge e, and let v_f be the initial vertex of e. Let r be the boundary label on f when read from v_f. Let $\gamma \subset \partial\mathcal{D}$ be a simple arc in the boundary ∂D from v to v_f, with label p.

Removing the interior of the edge e from $\mathcal{D}$ gives a van Kampen diagram $\mathcal{D}'$ with $M - 1$ bounded regions, possibly with some vertices of valence one. In the unbounded region of the diagram D', join to the base point v an arc labelled p leading to a disk with subdivided boundary labelled $r^{\pm 1}$. Continuing in this way, the whole diagram can be "deconstructed" to give a bouquet of circles describing w as a product $w = \prod_{i=1}^{M} p_i r_i {p_i}^{-1}$ (as in figure 2.3). Again the equality here is in the free group $F(\mathcal{A})$. Notice that the length of the words p_i is bounded by the number of edges in $\mathcal{D}$, which is bounded by $\ell(w) + M\rho$ where $\rho = \max_{r \in R} \ell(r)$. Alternatively, it is possible to find other words p_i of length bounded by $\ell(w)/2$ plus the length of the shortest path in D from the i-th face to the boundary. $\square$

It is important to underline the following aspects of the form of a van Kampen diagram. Regarding the diagram as a planar 2–complex, it is a collection of disjoint closed topological 2-cells joined by arcs (and vertices): removing the closures of these 2–cells from the diagram leaves a collection of trees. There is a retraction of the diagram onto a tree, realised by retracting each of the disc components to a point.

In fact the first part of the above proof establishes something stronger, where we may allow bounded regions of the planar graph to have labels which are not freely (nor cyclically) reduced. This will be useful when describing reduction of diagrams (see 2.7).

Proposition 2.4. *Let $\mathcal{D}$ be a finite, connected, oriented, based, labelled, planar graph where each oriented edge is labelled by an element of $\mathcal{A}$. Suppose in addition that the bounded regions of $\mathbb{R}^2 - D$ are labelled by words whose freely reduced forms* *lie in $\langle\langle R\rangle\rangle$, and that the boundary label is w. Then a finite, connected, oriented, based, labelled, planar graph D' can be obtained from D such that all bounded regions are labelled by cyclically reduced words in $\langle\langle R\rangle\rangle$, and the boundary label is the reduced word corresponding to w.*

Definition 2.5 (The presentation complex and the Cayley Complex)**.** Let $\mathcal{P} = \langle \mathcal{A} \mid R\rangle$ be a group presentation. The standard 2–complex $\mathcal{K}^2(\mathcal{P})$ associated to the presentation $\mathcal{P}$, consists of:

a single vertex v, and an oriented 1–cell ${e_i}^{(1)}$ for each $a_i \in \mathcal{A}$, labelled a_i;
a 2–cell ${e_j}^{(2)}$ for each $r_j \in R$.

The 1–cells are attached to the 0–skeleton in the only way possible. Thus the fundamental group of the 1–skeleton is the free group on the set $\mathcal{A}$. The 2–cell ${e_j}^{(2)}$ is attached to the 1–skeleton via a map identifying its boundary with a loop in the 1–skeleton corresponding to the word r_j. The resulting space is given the quotient topology. Giving all 1–cells unit length, and viewing the 2–cell ${e_j}^{(2)}$ as a regular euclidean polygon with $\ell(r_j)$ sides each of unit length, gives an additional piecewise euclidean structure to the complex. (When there are relations of length 1 or 2, use a disk with circumference length 1 or 2; such relators can of course be easily avoided.)

The Seifert–van Kampen theorem (in the simpler case when one component is simply connected) tells us that $\pi_1(\mathcal{K}^2(\mathcal{P}), v) = \Gamma(\mathcal{P})$.

The Cayley complex of the group $\Gamma(\mathcal{P})$ with respect to the presentation $\mathcal{P}$ is the universal covering space of $\mathcal{K}^2(\mathcal{P})$. This is obtained from $Cay^1(\Gamma, \mathcal{A})$ by adding 2–cells: for each $g \in \Gamma$, a 2–cell for each r_j is added, based at the vertex corresponding to g (i.e. a lift based at each vertex g of each attaching map). Notice that for a relation r, which is a proper power, say $r = s^q$, the vertex $g \in Cay^2(\Gamma, \mathcal{A})$ is in the boundary of q attached 2–cells labelled r, as there are q lifts of the attaching map which pass through this vertex, one based there, the others based at $gs^{q'}$ for $1 \leq q' < q$.

Examples:

1) The presentation $\langle x, y \mid xyx^{-1}y^{-1}\rangle$ gives a standard complex homeomorphic to the torus $S^1 \times S^1$, whose Cayley complex (as a P.E. complex) is (isometric to) the plane $\mathbb{R} \times \mathbb{R}$. The Cayley graph is the 1–skeleton, viewed as the set of points with at least one integer coordinate, and the vertices are those points with two integer coordinates.

2) The presentation $\mathcal{P} = \langle a, b \mid a^2, b^2\rangle$ gives a standard complex homeomorphic to two projective planes joined at a point. The Cayley complex $Cay^2(\mathcal{P})$ is homeomorphic to an infinite collection of spheres indexed by $\mathbb{Z}$, where the north pole of the i–th sphere is joined to the south pole of the $(i+1)$–st sphere.

Using this complex, we can give a topological derivation of van Kampen diagrams. This leads to diagrams which are essentially dual to van Kampen diagrams, and in their full generality are known as transversality diagrams. In this context they were introduced by Colin Rourke (see also Fenn's book [11]). As is customary, we take for granted many results about transversality. The advantage of this alternative treatment is that the diagrammatic method is based on topology rather than combinatorics, though this is perhaps mostly a matter of taste.

Definition 2.6 (Pictures)**.** Consider the presentation $\mathcal{P} = \langle \mathcal{A} \mid R \rangle$, the standard complex $\mathcal{K}^2(\mathcal{P})$ and $w \in \langle\langle R \rangle\rangle$. Then w defines a path γ_w in the 1–skeleton $\mathcal{K}^2(\mathcal{P})^{(1)}$ which is null–homotopic in $\mathcal{K}^2(\mathcal{P})$. This means that there is a map $f : (D, \partial D) \to (\mathcal{K}^2(\mathcal{P}), \mathcal{K}^2(\mathcal{P})^{(1)})$ such that $f|_{\partial D} = \gamma_w$, and the homotopy lifting property says that f lifts to a map $(D, \partial D) \to (Cay^2(\mathcal{P}), Cay^1(\mathcal{P}))$.

• After a homotopy (relative to ∂D), we can suppose that, for each relator r_j, f is transverse to the centre $\hat{e}_j$ of the corresponding 2–cell e_j. This means that $f^{-1}(\hat{e}_j)$ is a finite set of points in $intD$, and that there is a disjoint set of open neighbourhoods $V(\alpha_{j,k})$ of the points $\alpha_{j,k} \in f^{-1}(\hat{e}_j)$ such that the restriction of f to each neighbourhood is a homeomorphism into a neighbourhood of $\hat{e}_j$ in $int(e_j)$.

• As $f(D - \cup_{j,k} V(\alpha_{j,k})) \subset \mathcal{K}^2(\mathcal{P}) - \cup_j \hat{e}_j$, and this latter space retracts onto the 1–skeleton $\mathcal{K}^2(\mathcal{P})^{(1)}$, after a further homotopy of f (fixing the boundary ∂D), we can suppose that $f(D - \cup_{j,k} V(\alpha_{j,k})) \subset \mathcal{K}^2(\mathcal{P})^{(1)}$.

• Now, by a further homotopy (fixing the boundary), we can make $f : D - \cup_{j,k} V(\alpha_{j,k}) \to \mathcal{K}^2(\mathcal{P})^{(1)}$ transverse to the mid–points $\hat{e}_i$ of the 1–cells ${e_i}^{(1)}$. This means that $f^{-1}(\hat{e}_i)$ is a finite set of properly embedded arcs and loops in $D - \cup_{j,k} V(\alpha_{j,k})$. Moreover each arc/loop has a transverse orientation and label coming from the orientation and label in a neighbourhood of $\hat{e}_i$ in ${e_i}^{(1)}$. (Here we should note that during the construction of the complex $K(\mathcal{P})$, the attaching maps of the 2–cells should have been made transverse to the points $\hat{e}_i$.)

The *picture* corresponding to f is the disc D together with the collection of subdisks (or "fat vertices") $V_{j,k}$ and the embedded loops and arcs. Each arc and loop is transversely oriented and labelled by some $x \in \mathcal{A}$, inducing labels r_j on the boundary of each $V_{j,k}$, and a label w on ∂D, when read from appropriate points, and in an appropriate direction.

The picture can be thought of as constructed from copies of small discs ("fat vertices") with protruding "legs", each of which has a transverse orientation and a label from $\mathcal{A}$: each disc corresponds to some $r \in R$, and the labels on the edges, read from some base point, with exponents ± 1 according to the orientation spell out the word r (Roger Fenn calls these "spiders and anti–spiders" in [11]).

Notice that there may be free loops in the picture: i.e. there may be a simple loop in D which maps to a point $\hat{e}_i^{(1)}$. Such a loop, and all of the picture in the subdisk of D bounded by it, can be removed to obtain a simpler picture. In fact the interior of the subdisk corresponds to a picture on a 2–sphere.

It is easy to see that pictures are basically dual to diagrams: to obtain a picture from a diagram: surround the diagram in the plane by a big circle. Insert

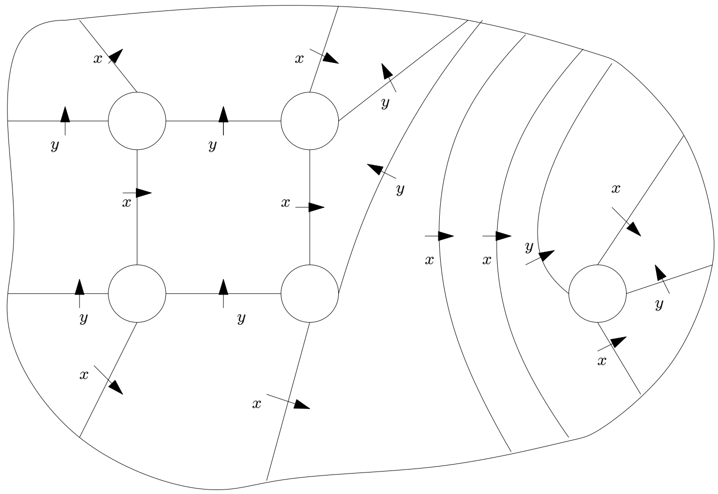


Figure 2.8: A picture for $w = x^3yx^{-1}y^{-1}x^{-2}y^2x^{-2}y^{-2}x^2$, *cfr* 2.1

a small circle in the interior of each compact region. Dual to an edge separating two faces of the diagram, insert an edge joining the added circles in each face. Label and transversely orient these added edges according to the orientation of the original edge.

Given a picture, remove all free loops. Around each small disc with n legs, draw a polygon with n sides. The sides are labelled and oriented according to the label and (transverse) orientation of the legs. For each arc in the picture, the sides of polygons occurring at the two ends of the arc are identified: the danger here is that this may lead to a non–planer diagram (thus the word "basically" above).

There is an obvious simplification that can performed on diagrams and on pictures:

Definition 2.7. Let D be a van Kampen diagram. Let F_1 and F_2 be distinct compact regions of D, such that there is at least one edge e in the intersection $\partial F_1 \cap \partial F_2$.

For $i = 1, 2$, let r_i be the label read on the boundary ∂F_i starting from the initial vertex of e, and reading in the direction induced by the orientation of e (see the example figure 2.10). If $r_1 = r_2$ (i.e. identical as words in $F(\mathcal{A})$) then the diagram is said to be *unreduced*, and if no such pair of faces exists then the diagram is said to be reduced.

Removing the edge e from the unreduced diagram gives the possibility of performing a series of foldings which identify the rest of the boundaries of F_1 and F_2, while leaving unaffected the remainder of the diagram (as in the proposition 2.4). In this way a diagram with two fewer regions is obtained (maybe there are

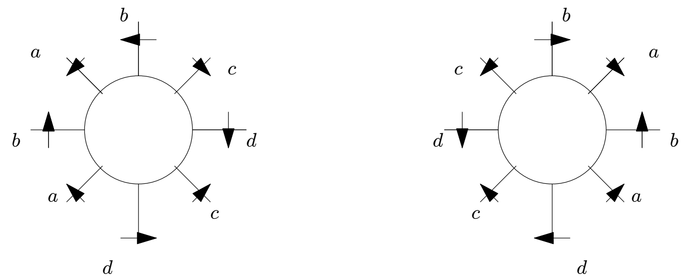


Figure 2.9: A spider and an anti–spider for the relation $r = aba^{-1}b^{-1}cdc^{-1}d^{-1}$

many fewer regions after the folding has finished). As an exercise in this dual method, we now describe the reduction procedure in the world of pictures.

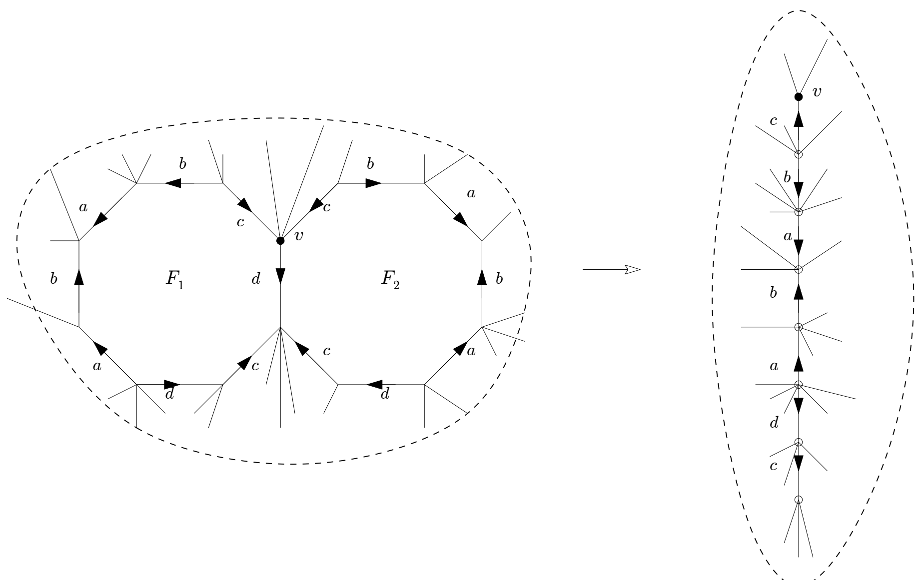


Figure 2.10: A reduction in a van Kampen diagram; here the common edge e, labelled d, is removed, then the other edges are in turn identified.

Translating into the world of pictures, let P be a picture in which two small discs are joined by at least one arc, and such that the labels on the two small discs, starting from the label on one particular arc joining them, and in the direction induced by the transverse orientation on that arc, are the same. Then the picture can be altered to remove these two small disks as indicated in figure 2.11.

We saw in the world of van Kampen diagrams that care has to be taken when

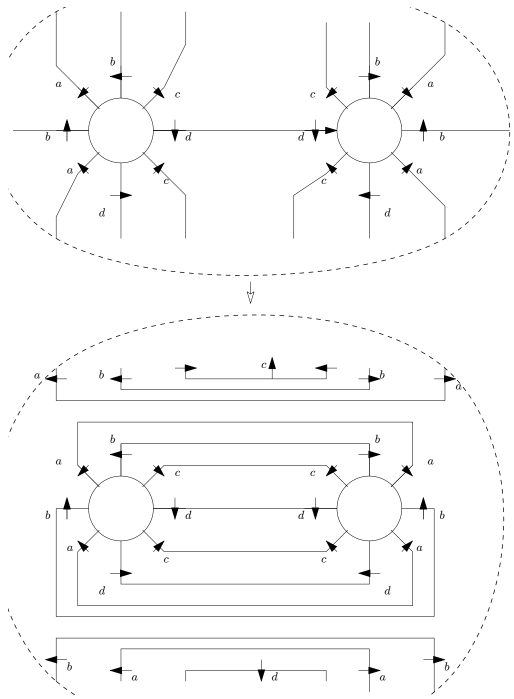


Figure 2.11: A reduction of cancelling spider and anti–spider

there is more than one segment in the intersection of the boundaries of F_1 and F_2, and in this case part of the diagram may have to be discarded in order to retain planarity. This problem translates in the pictures world to two collections of arcs joining the two small disks with possibly some subpicture trapped between them as in figure 2.12. The reduction process can then lead to a subpicture unconnected with the rest of the picture. This subpicture can be discarded without affecting the boundary label, or the essential properties of the picture.

There is much which can be said about reduced diagrams and the relationship with π_2 and the various definitions of “aspherical”. In this topological version, it is clear that a map of a sphere into the Cayley complex leads to a picture on a sphere, using the same transversality and homotopy construction as was used for the map of a disc into the complex. (See [8] for a detailed discussion of some of the issues relating to this method of studying π_2, including the problems with the various definitions of *aspherical* used in [20]).

We have noted that when $w =_{\mathcal{P}} 1$ there is a diagram for w over $\mathcal{P}$, and there is clearly a reduced diagram, and in fact a smallest diagram, with the smallest number of compact faces, i.e. a diagram D such that $area(D) = area(w)$ is reduced. The area of a picture is naturally the number of small disks (fat vertices).

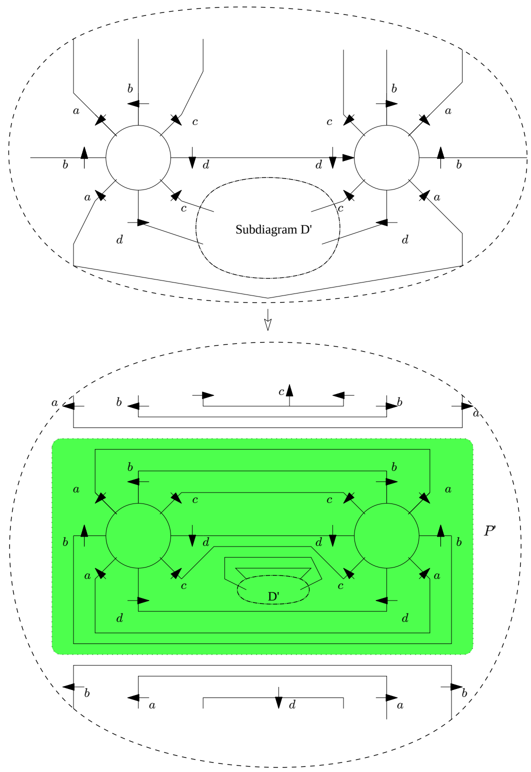


Figure 2.12: Reduction produces a disconnected picture: the subpicture P' can then be discarded.

There are certain presentations where for each word w representing the trivial element, there is a unique reduced diagram.

Britton's Lemma and Collins' Lemma for HNN extensions via pictures

As usual, let $\langle \mathcal{A} \mid R\rangle$ be a presentation of a group Γ. Consider the HNN extension :

$H = \Gamma *_C = \langle \Gamma, t \mid tct^{-1} = \phi(c), c \in C\rangle = \langle \mathcal{A}, t \mid R, tct^{-1} = \phi(c), c \in C\rangle$

for some subgroup $C < \Gamma$ and some injective homomorphism $\phi : C \to \Gamma$. To ensure that this group is finitely presented, we suppose that $\mathcal{A}, R$ are finite, and that C is finitely generated.

Lemma 2.8 (Britton's Lemma)**.** *Let $w = b_1 t^{\alpha_1} b_2 t^{\alpha_2} \dots b_k t^{\alpha_k}$ be a non–empty reduced word in $F(\mathcal{A}, t)$, where $b_i \in F(\mathcal{A})$, such that $\alpha_i \neq 0$ for $i < k$.*

If $w =_H 1$ and $\alpha_1 \neq 1$, then for some $i = 2, \dots, k$, either $\alpha_{i-1} > 0 > \alpha_i$ and $b_i \in C$, or $\alpha_{i-1} < 0 < \alpha_i$ and $b_i \in \phi(C)$.

(There are two such subwords b_i if sub–indices are considered modulo k .)

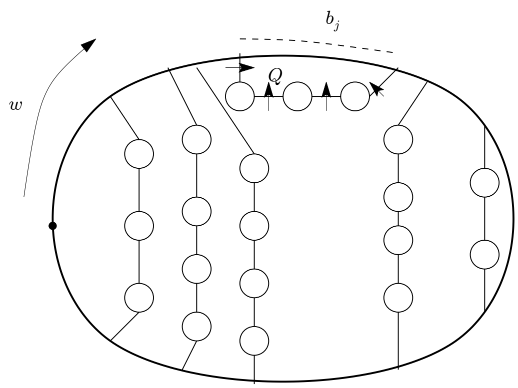


Figure 2.13: The proof of Britton's Lemma: suppress all edges except those labelled by t, and all relation disks except those labelled by $tat^{-1}\phi(a)^{-1}$. If there are rings formed by the t–edges then these can be removed as in the proof of 2.9 below. There is some "outermost region" bounded by t edges (here labelled Q) which does not contain the base point v, and this region meets the boundary in a segment labelled by the subword b_i of w. The labels on the two t–edges at either end of the b_j segment are opposite due to the consistency of the t orientation along the chain of t edges and relation discs. Thus $t^\epsilon b_i t^{-\epsilon}$ is a subword of w.

Once this result has been established, we can assume that elements in H are represented by words which do not contain "pinches". A *pinch* in a word on the generators for the HNN extension H is a subword of the form tbt^{-1} where $b \in C$, or of the form $t^{-1}bt$, where $b \in \phi(C)$ (a cyclic pinch is a pinch in the word viewed cyclically). Detecting the existence of such pinches in w depends upon having an algorithm to decide whether or not a word in the generators for H represents an element of $C \cup \phi(C)$. The problem of deciding, for a finite set of elements X in the group G, whether or not a word w lies in the subgroup generated by X, is known as the *generalised word problem*. A consequence of Britton's lemma is thus:

Proposition 2.9 (The word problem for HNN extensions)**.** *The word problem is solvable for H if the word problem is solvable for Γ and the generalised word problem for C and for $\phi(C)$ in Γ is solvable.*

Proof. Notice that the solution of word problem for Γ is included in the solution of the generalised word problem for C. Let $w = b_1 t^{\alpha_1} b_2 t^{\alpha_2} \dots b_k t^{\alpha_k}$ be a cyclically reduced word in $F(\mathcal{A}, t)$.
Case 1: there are no occurrences of t in w.
If $w =_H 1$, then there is a corresponding picture in which no t–edges meet the boundary. If there are no t–edges in the picture, then $w = 1$ in Γ, and the algorithm for the word problem in Γ tells us whether or not $w = 1$ in Γ. If there are t–edges in the picture, then they are joined to relation discs in such a way as to form rings. An innermost such ring has a label w' in $F(\mathcal{A})$, and the sub–picture enclosed is a picture for $w' = 1$ in Γ. The word on the outside of this ring is either $\phi(w')$ or $\phi^{-1}(w)$, according to the orientation of the t-edges. In either case, $w' = 1$ in Γ if and only if $w' \in C$ and $\phi(w') = 1$ in Γ, or $w' \in \phi(C)$ and $\phi^{-1}(w') = 1$ in Γ. In both cases, an innermost such ring can be removed and replaced by a picture over Γ for the outside word ($\phi(w')$ or $\phi^{-1}(w')$) with no t–edges. In this way, after repetitions, if $w = 1$ in H and there are no t occurrences in w, then there is a picture for w over Γ (and thus the natural homomorphism of Γ into H induced by the map on the generators is in fact injective).

Case 2: There are occurrences of t in w: proceed by induction on the number of such t–occurrences.
Suppose there is an algorithm to decide triviality when a word has at most N t–occurrences. Suppose that w is a word with $N+1$ t–occurrences. If $w = 1$ in H, then Britton's Lemma says that there is a subword of w of the form tb_jt^{-1} or $t^{-1}b_jt$ for some j, such that $b_j \in C$ or $b_j \in \phi(C)$ respectively. The algorithm to decide membership of the subgroups C and $\phi(C)$ in Γ decides whether or not such subwords exist in w, and when such a subword is found, it is replaced by $\phi(b_j)$ or $\phi^{-1}(b_j)$ respectively, giving a new word with fewer t–occurrences which is equal to w in H. Notice that here $\phi(b_j)$ is calculated via a semigroup homomorphism when C is given as a finitely generated subgroup of Γ, and $\phi(c_i)$ is given for each generator of C. □

The conjugacy problem for HNN extensions can be studied in a similar way. This diagrammatic proof was given by Miller and Schupp [22]. Here the relevant lemma concerns annular pictures:

Lemma 2.10 (Collins' Lemma)**.** *Let $u = b_1 t^{\beta_1} b_2 t^{\beta_2} \dots b_p t^{\beta_p}$, $v = d_1 t^{\delta_1} d_2 t^{\delta_2} \dots d_q t^{\delta_q}$ be cyclically reduced words in $F(\mathcal{A}, t)$ such that $b_i, d_j \in F(\mathcal{A}) - \{1\}$, and $\beta_i, \delta_j \in \mathbb{Z} - \{0\}$, except when $u = b_1$ or $v = d_1$. Suppose in addition that there are no pinches or cyclic pinches in u or v.*
If u and v are conjugate in $H = \langle \Gamma, t \mid tCt^{-1} = \phi(C) \rangle$, then one of the following holds:
- *u, v are words in $F(\mathcal{A})$ which are conjugate in Γ;*
- *there is finite chain of words in $F(\mathcal{A})$, $u = v_0, u_1, v_1, u_2, v_2, \dots, u_k, v_k, u_{k+1}$ such that $v_i = \phi^{\pm 1}(u_i)$, as group elements, $u_i, v_i \in C \cup \phi(C)$, and for each $i = 0, \dots k$, v_i is conjugate to u_{i+1} by an element of Γ.*
- *both u and v contain t occurrences, and $p = q$, $u = b_1 t^{\beta_1} b_2 t^{\beta_2} \dots b_p t^{\beta_p}$ is conjugate to some cyclic conjugate of v of the form $d_i t^{\delta_i} d_{i+1}^{\delta_{i+1}} \dots d_q t^{\delta_q} d_1 t^{\delta_1} \dots d_{i-1} t^{\delta_{i-1}}$ and the conjugation can be realised by an element of $C \cup \phi(C)$.*

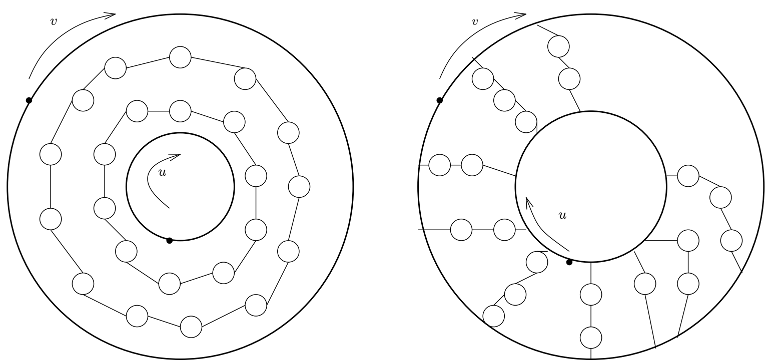


Figure 2.14: The two possible non–trivial conjugacy pictures for Collins' Lemma are annular pictures where the boundary components are labelled u and v (the first case of the statement corresponds to a picture without t–edges). Again suppress all edges except for t–edges, and suppose that there are some t–occurrences in u and/or v. Remove closed loops of t–edges which bound discs in the annulus as in the proof of 2.9 above. As there are no cyclic pinches, all chains of t–edges and relations go from one boundary component to the other, or form closed loops which are "parallel" to the boundary (i.e. go once round the annulus). These two types cannot coexist, and this leads to the second two cases of the statement.

The conjugacy problem is in general not solvable for HNN extensions, even when the conjugacy and word problems are solvable in Γ. The difficulty lies in the estimation of the number of conjugations in the second case of Collins' lemma.

If the generalized word problem is solvable for C and $\phi(C)$ in Γ, then cyclic pinches can be detected in words in $F(\mathcal{A}, t)$.

If after removal of these pinches from two words one of the words contains t–occurrences, then the other word must also have t–occurrences, and case 3 of Collins' Lemma applies.

Chapter 3

Small cancellation conditions

Certain conditions on the form of the relators in a presentation, and how they interact locally, can give strong restrictions on the properties of the group so presented. The so–called "small cancellation conditions" developed in the sixties by Lyndon, Greendlinger and others quickly give isoperimetric inequalities. A full presentation is given in chapter V of Lyndon and Schupp's book [20]. These conditions are also sufficiently generic to give results about "random" presentations (see for instance [25]). The idea is to use elementary properties of graphs in the plane, in particular the Euler formula: $V - E + F = 2$. That is, given a non–empty finite connected planar graph, the sum of V, the number of vertices, and F, the number of components of the complement, minus E, the number of edges, is equal to 2.

Notice that in a van Kampen diagram over a finite presentation $\mathcal{P}$, when the boundaries of two compact faces have a common arc of intersection, the corresponding relators (read cyclically) contain a common subword. If these common subwords are always very short, then a face of the diagram which does not meet the boundary meets many other faces. Given an arc common to the boundaries of two faces, vertices of degree 2 can be suppressed and the label changed to the corresponding word in the generators. If all vertices of degree 2 are suppressed in this way, the planar graph now has all vertices of degree at least 3. A consequence of the Euler formula is that in a planar graph in which all vertices have degree at least three, there are at least two compact regions with at most 5 sides, as can be seen as follows:

If all vertices have degree at least 3, counting the edges at each vertex gives $2E \geq 3V$ and $V \leq 2E/3$ (as each vertex meets at least three edges, and each edge is counted twice).

If all faces except two (for instance the unbounded region and one compact region) have at least 6 sides (the other two having at least 1 side each), then $E \geq 6(F-2)/2 + 2/2$ implies that $F \leq E/3 + 5/3$ (as each of $F-2$ faces meets at least six edges, and each edge is counted twice).

Thus $V + F - E \leq E - E + 5/3 = 5/3$ and is therefore not 2.

Thus there at least three regions with 5 or fewer edges.

In order to apply this to van Kampen diagrams, we need to slightly alter our view of the diagrams, and for this we need the following definitions.

Definition 3.1 (The small cancellation conditions)**.** Let $\mathcal{P} = \langle \mathcal{A} \mid R \rangle$ be a finite presentation. Let R^C be the set of all reduced cyclic conjugates of elements of R and their inverses.

A non–trivial word $p = a_1 a_2 \dots a_k \in F(\mathcal{A})$ is a *piece* with respect to R if there are two different relations $r_1, r_2 \in R^C$ such that $r_1 = ps_1$ and $r_2 = ps_2$ (and of course $s_1 \neq s_2$). That is, p occurs as a subword of two different relations in R^C. Notice that non–trivial subwords of pieces are also pieces.

The presentation satisfies the condition :

• $C'(1/\lambda)$ for $\lambda \in \mathbb{R}^+$ if for all pieces p, if p is a subword of $r \in R^C$, then $\ell(p)/\ell(r) < 1/\lambda$.

• $C(k)$ if for all relations $r \in R^C$, it is not possible to write r as a product of fewer than k pieces.

Clearly the property $C'(1/\lambda)$ implies the property $C'(1/\lambda')$ if $\lambda \geq \lambda'$, and $C(k)$ implies $C(k')$ if $k' \leq k$.

The first non–trivial example to consider is the fundamental group of a genus two closed orientable surface, with the standard presentation $\langle a, b, c, d \mid aba^{-1}b^{-1}cdc^{-1}d^{-1} \rangle$. Here each generator is a piece, and no word of length 2 is a piece. Thus the presentation satisfies the conditions $C'(1/7)$ and $C(8)$.

Notice that for positive integers k, the metric condition $C'(1/k)$ implies the condition $C(k+1)$. For a presentation which satisfies the condition $C(k)$, in all <u>reduced</u> van Kampen diagrams, each compact face not meeting the boundary has at least k sides (when degree 2 vertices are suppressed). This is because, when a long subword labels a segment separating two faces of a diagram, the relations must be equal in F (and so the same element of R^C), when read from the initial point of the common segment. It follows that the diagram is not reduced.

Lemma 3.2 (The small cancellation lemma : non–metric version)**.** *Let Δ be a van Kampen diagram which is a topological disk, containing at least 2 regions, such that all internal regions have ≥ 6 sides, and all vertices have degree ≥ 3. For $i = 1, \dots, 5$, let b_i be the number of regions meeting $\partial\Delta$ in exactly one connected segment (and possibly some vertices) having exactly i internal edges, forming a connected part of their (internal) boundary.*

Then $3b_1 + 2b_2 + b_3 \geq 6$.

One of the main applications of this important lemma is:

Theorem 3.3. *If a finite presentation $\mathcal{P}$ satisfies the condition $C'(1/6)$, then Dehn's algorithm solves the word problem.*

First we prove the theorem, supposing that the lemma has been proved.

Proof of the theorem. First note that $\mathcal{P}$ also satisfies the non–metric $C(7)$ (and so the $C(6)$) condition, and that the metric condition implies that the sum of the lengths of 3 pieces of a relator r is strictly less than half the length of r.

Let w be a cyclically reduced non–trivial word representing the identity element of the group, and let D be a diagram for w over $\mathcal{P}$. We show that w contains a subword which is more than half of a relator.

The form of the diagram is that of a collection of topological discs in the plane joined by trees, such that regarding the topological discs as fat vertices, the whole is a tree. As w is supposed to be cyclically reduced, there are no vertices of degree 1 to be suppressed (removing them and the edges meeting them would change w by a cyclic reduction). Either the diagram now consists of a single topological disc, or there are at least two extremal topological discs which intersect the rest of the diagram in just one vertex.

Case 1: D consists of a single topological disc. The lemma provides two faces F_1, F_2 meeting the boundary ∂D in segments of length $> \ell(F_1)/2$ and $> \ell(F_2)/2$. This means that the segment corresponding to $\partial D \cap \partial F_1$ is labelled by a subword u of some r, a cyclic conjugate of a relator (or its inverse) in R, of length greater than $\ell(r)/2$, as it is all of r except for at most 3 pieces. Thus $r = uv$ such that $\ell(u) > \ell(v)$, and so $u =_\Gamma v^{-1}$ and $w = w'uw''$ and $w =_\Gamma w'vw''$, and $\ell(w'vw'') < \ell(w'uw'')$.

Case 2: D contains more than one topological disc. There are at least two disc components which are extremal, i.e. which meet the rest of the diagram in a single vertex. In each of these components, at least one of the two boundary segments given by the lemma has an interior which does not contain this vertex. Thus the diagram contains two faces meeting the boundary in the manner claimed.

□

Proof of the lemma. The following proof is given by L.I. Greendlinger and M.D. Greendlinger in [16]. Let Δ be a reduced van Kampen diagram over the presentation $\mathcal{P}$ which is a topological disc, and suppose that Δ satisfies the $C(6)$ condition.

Create a diagram $\mathcal{G}$ on S^2 by doubling Δ along $\partial\Delta$, after a twist by less than half the length of the shortest edge on $\partial\Delta$. This gives a graph on S^2.

The twist creates a new vertex in each edge of $\partial\Delta$. Each region in Δ contributing to b_i has $i+1$ sides in its boundary, and the region intersects the interior of Δ in a consecutive series of i of these edges (and possibly some other vertices). In this way, every region of Δ contributing to b_1, b_2, b_3 gives 2 regions in $\mathcal{G}$, each with 3,4, and 5 sides respectively.

All other bounded regions of Δ give 2 regions in $\mathcal{G}$ each with at least 6 sides (including regions meeting the boundary in more than one connected segment).

Let V, E, F denote the numbers of vertices, edges and regions of $\mathcal{G}$.

$E \geq (6(F - 2b_1 - 2b_2 - 2b_3) + 3.2b_1 + 4.2b_2 + 5.2b_3)/2$

$\implies E \geq 3F - 3b_1 - 2b_2 - b_3 \implies F \leq E/3 + (3b_1 + 2b_2 + b_3)/3$
Also $E \geq 3V/2$ implies $V \leq 2E/3$, so
$2 = V - E + F \leq 2E/3 - E + E/3 + (3b_1 + 2b_2 + b_3)/3$
which in turn implies that $3b_1 + 2b_2 + b_3 \geq 6$. □

It is not hard to give a second condition on the combinatorics of a presentation which guarantees that internal vertices have degree at least 4:

Definition 3.4. Let $\mathcal{P} = \langle \mathcal{A} \mid R \rangle$ be a finite presentation. Let R^C be the set of all reduced cyclic conjugates of elements of R and their inverses. The presentation satisfies the condition $T(q)$ if :
for all $3 \leq k < q$, for all $r_1, r_2, \ldots, r_k \in R^C$, and for all pieces $p_1, \ldots, p_k$, if $r_1 = p_1 {r_1}' {p_2}^{-1}$, $r_2 = p_2 {r_2}' {p_3}^{-1}$, $\ldots$,, and $r_k = p_k {r_k}' {p_1}^{-1}$, then for some $i \mod k$, it is the case that $r_i = {r_{i+1}}^{-1}$.

This complicated looking definition just states that in a reduced diagram, all internal vertices have degree at least q.

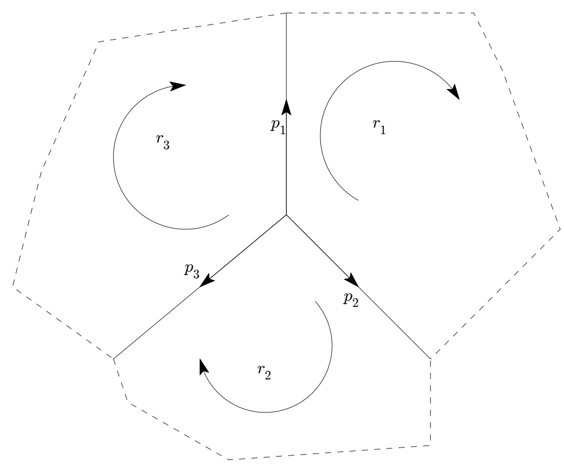


Figure 3.1: The $T(q)$ condition with $k = 3$.

With this definition, the same methods as above can be used to show :

Lemma 3.5 (The small cancellation lemma $C(4) - T(4)$ version)**.** *Let $\mathcal{P}$ be a finite presentation satisfying the conditions $C(4){-}T(4)$, and D be a reduced van Kampen diagram over $\mathcal{P}$ which is a topological disc containing more than one face.*

Then D contains two bounded faces $\partial F_1, \partial F_2$, such that for $i = 1, 2$, $\partial F_i \cap \partial D$ contains a connected segment containing all of ∂F_i except for at most 2 pieces.

Theorem 3.6. *A presentation which satisfies $C(6)$ or $C(4) - T(4)$, satisfies a quadratic isoperimetric inequality.*

Proof. We consider a van Kampen diagram Δ satisfying the $C(6)$ condition, where vertices of degree 2 have been suppressed.

We first consider the case when Δ is a topological disk.

Note that if F has a non–boundary edge, both of its endpoints lying on the boundary, splitting Δ into two subdiagrams Δ_1, Δ_2, each with $\delta_i \geq 4$ vertices of degree at least 3 on its boundary, then $\delta_1 + \delta_2 + 2 \geq \delta$. The induction hypothesis, that the number of faces is bounded by the square of the number of vertices on the boundary of a topological disc diagram (when vertices of degree 2 are suppressed) tells us that $F = F_1 + F_2 \leq {\delta_i}^2 + {\delta_2}^2 \leq (\delta_1 + \delta_2 - 2)^2$ if $\delta_1, \delta_2 \geq 4$.

We now suppose that there are no such edges with both endpoints on the boundary. Thus there any region of Δ meeting $\partial\Delta$ in at least two segments has at least 6 sides in Δ.

We use the notation of 3.2, and in addition we denote by b_5' the sum of $\sum_{i>4} b_i$ and the number of regions of Δ meeting $\partial\Delta$ in at least one segment and having at least 5 sides, and we denote by V' number of vertices in Δ having degree at least 4.

Notice that the regions contributing to b_4 give rise to regions in $\mathcal{G}$ with 6 sides, while those contributing to b_5' give rise to regions with at least 7 sides.

Counting the edges in $\mathcal{G}$ via the faces, we get

$$\begin{aligned} E &\geq 3(F - 2(b_1 + b_2 + b_3 + b_4 - b_5')) + 3b_1 + 4b_2 + 5b_3 + 6b_4 + 7b_5' \\ &\geq 3F - 3b_1 - 2b_2 - b_3 + b_5' \\ \implies F &\leq E/3 - (3b_1 + 2b_2 + b_3 - b_5')/3 \end{aligned}$$

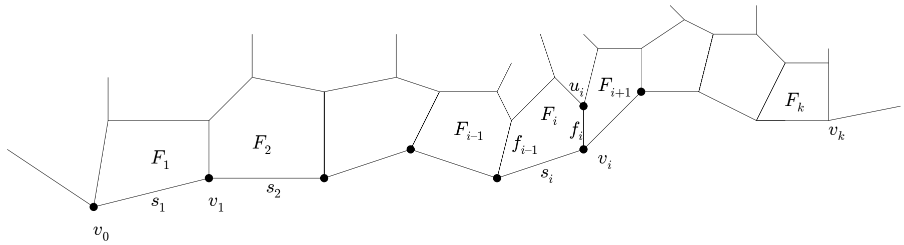
Estimating the number of edges of $\mathcal{G}$ via the vertices, we get

$$\begin{aligned} E \geq (3(V - 2V') + 4.2V')/2 = 3V/2 + V' \\ \implies V \leq (2E - 2V')/3 \end{aligned}$$

Applying these inequalities to the Euler characteristic of the graph $\mathcal{G}$ on the sphere, we get

$$\begin{aligned} 2 = V - E + F &\leq (2E - 2V')/3 - E + E/3 - (3b_1 + 2b_2 + b_3 - b_5')/3 \\ \implies 6 &\leq (3b_1 + 2b_2 + b_3) - (2V' + b_5') \end{aligned}$$

It follows that there are at least two more regions meeting the boundary of Δ contributing to $(3b_1 + 2b_2 + b_3)$ than there are regions and vertices contributing to $(2V' + b_5')$. If, between each pair of vertices contributing to $3b_1 + 2b_2 + b_3$, there is a boundary vertex contributing to $(2V' + b_5')$, then we would have $2V' + b_5' \geq V' \geq 3b_1 + 2b_2 + b_3$. From this we deduce that there is a segment s on $\partial\Delta$ meeting faces $F_1, F_2, \ldots, F_k$ in subsegments $s_1, s_2, \ldots, s_k$ such that



- $\partial F_i \cap S = s_i$;
- $s_i \cap s_{i+1} = v_i$, a vertex in $s \in \partial\Delta$ of degree 3;
- F_1 and F_k contribute to $3b_1 + 2b_2 + b_3$: have boundary consisting of one edge in $s \subset \partial\Delta$ together with a sequence of at most 3 interior edges;
- for $i = 2, \ldots, k-1$, each region F_i contributes to b_4, and it has a boundary consisting of one edge in $s \subset \partial\Delta$ together with a sequence of at most 3 interior edges;
- for $i = 2, \ldots, k-1$, each intersection $F_i \cap F_{i+1}$ contains an edge which meets the vertex v_i, the other end being the vertex u_i.

Also, if the first vertex v_0 has degree 4, then it will contribute to $2V'$, though it may be counted twice if it is also the final vertex of another chain. In any case, we may assume that v_0 and v_k, the final vertex of F_k are of degree 3.

We show that for this (topological disk diagram) Δ, the number of faces is bounded above by the square of δ, the number of vertices in $\partial\Delta$.

Consider the edge f_i of the chain: the vertices of this edge are $v_i = f_i \cap s$ and $u_i \notin \partial\Delta$. Suppose that one of the other edges at u_i, say e, has an endpoint on $\partial\Delta$ (and so u_i has degree at least 4). Then cutting Δ along the edges f_i and e splits Δ into two diagrams as before, each of which has at least 4 vertices on its boundary, and as before the induction hypothesis applies to show that $F \leq \delta^2$.

We can now suppose that none of the vertices u_i associated to the chain s is joined to $\partial\Delta$ (other than by f_i). If in each such chain there were a vertex u_i of degree at least 4, then again $V' > 3b_1 + 2b_2 + b_3$. Thus we can suppose that each vertex u_i has degree 3.

Removing the faces $F_1, \ldots, F_k$ and the edges $s_1, \ldots s_k, f_1, \ldots, f_{k-1}$, from Δ gives a diagram Δ' which is a topological disc with k fewer faces. The vertices v_0 and v_k have degree 2, so that after their suppression, and the suppression of the vertices u_i, all now of degree 2, the number δ' of vertices in boundary of Δ' is at most $\delta - 2$. Thus $F = F' + k \leq (\delta - 2)^2 + k \leq \delta^2 - 2\delta + 4 + k$. But $k \leq \delta$ and $\delta \geq 4$ so that $F \leq \delta^2$.

Now that the quadratic inequality has been established for topological disk diagrams without vertices of degree 2, it suffices to notice that a general van

Kampen diagram without vertices of degree 2 satisfies the same inequality, as such a diagram is made up of several disc diagrams, the area is the sum of the areas of the disc components, and the length of the boundary is at least the sum of the boundary lengths.

Finally, to recover the actual isoperimetric inequality for the presentation, the vertices of degree 2 must be re–instated. But on any edge in the boundary of a region corresponding to the relation r, there at most $\max \ell(r)$ vertices.

Thus $Area(w) \leq \max \ell(r)\ell(w)^2$.

The result for $C(4) - T(4)$ presentations follows in the same way. $\square$

The same sort of proof can be used to solve conjugacy problems, using diagrams on an annulus rather than a disk. One application of this is Weinbaum's solution [30] of the conjugacy problem for alternating knot groups (see also [20, Chapter V]). The first step in the proof is to show that a prime tame alternating knot k in S^3 gives rise to a presentation for $\pi_1(S^3 - k) * \mathbb{Z}$ which satisfies the conditions $C(4) - T(4)$.

Chapter 4

Isoperimetric inequalities and quasi–isometries

As we have stated earlier, solving the word problem for a presentation $\mathcal{P} = \langle \mathcal{A} \mid R \rangle$ involves showing that an expression for w as a product of conjugates of relators $w = \prod_{i=1}^{M} p_i r_i^{\epsilon_i} p_i{}^{-1}$ can be found, with $p_i \in F(\mathcal{A})$, $r_i \in R$, and $\epsilon_i = \pm 1$. The smallest such number M is called the area of w. The function $\delta_{\mathcal{P}} : \mathbb{N} \to \mathbb{N} : \delta_{\mathcal{P}}(n) = \max\{Area(w) \mid w \in \langle\langle R \rangle\rangle, \ell(w) \leq n\}$ is called the Dehn function of the presentation. An *isoperimetric inequality* for the presentation is a function $f : \mathbb{N} \to \mathbb{R}$ such that for all $n \in \mathbb{N}$, $\delta_{\mathcal{P}}(n) \leq f(n)$.

Geometric group theory is concerned with properties of groups which are invariant under quasi–isometries (this is almost a definition of the theory):

Definition 4.1. Let $\lambda \geq 1$ and $\epsilon \geq 0$ be constants, and let X, Y be metric spaces.

A map $f : X \to Y$ is a (λ, ϵ)*–quasi–isometry*, or a (λ, ϵ)*quasi–isometric embedding*, if for every pair of points $x, x' \in X$ we have

$$\frac{1}{\lambda} d_X(x, x') - \epsilon \leq d_Y(f(x), f(x')) \leq \lambda d_X(x, x') + \epsilon.$$

If there is such a (λ, ϵ)–quasi–isometry and a constant C such that for all $y \in Y$, there is a point $x(y) \in X$ such that $d_Y(f(x(y)), y) < C$ (i.e. if f is 'almost surjective') then we say that X and Y are quasi–isometric. The function $Y \to X$ given by $y \to x(y)$ is a *quasi–inverse* of the function f. In general a map $g : Y \to X$ is a quasi–inverse of the map f if there is a constant C such that $\forall x \in X, d_X(x, g \cdot f(x)) < C$ and $\forall y \in Y, d_Y(y, f \cdot g(y)) < C$.

Warning: care must be taken here as a quasi–isometry $f : X \to Y$ in general is far from being a surjection: for instance the inclusion of $\mathbb{R}$ in $\mathbb{R} \times \mathbb{R}$ as a factor is a quasi–isometry (a quasi–isometric embedding).

The standard examples are:

Bounded spaces are quasi–isometric to a point;

The inclusion $\mathbb{Z} \hookrightarrow \mathbb{R}$ is a quasi–isometry, and in fact is almost onto, with the function "integer part" providing a quasi–inverse.

The inclusion $\mathbb{Z} \times \mathbb{Z} \hookrightarrow \mathbb{R} \times \mathbb{R}$, can be viewed as a quasi–isometric embedding of a Cayley graph of the fundamental group of a torus in the universal covering space, and this inclusion is almost onto.

In the same way, the tessellation of the hyperbolic plane $\mathbb{H}^2$ by regular octagons with corner angle $\pi/4$ describes the universal covering space of a surface S_2 of genus 2. The dual graph can be viewed as a quasi–isometric embedding of the Cayley graph of the fundamental group $\pi_1(S_2)$ in $\mathbb{H}^2$. This is a special case of the result due to Švarc [28] and to Milnor [23] that the universal covering space of a closed compact Riemannian manifold is quasi–isometric to the Cayley graph of the fundamental group. We shall prove a more general form of the statement, following Bridson and Haefliger's approach [6, p.140]. In order to state the result in a little more generality, we introduce the following definitions: a metric space (X, d) is *proper* if closed balls $\overline{B}_r(x)$ are compact, and is a *length space* if for all points $x, y \in X$, $d(x, y) = \inf \ell(\gamma)$, where the infimum is taken over all rectifiable curves in X from x to y. Recall that a curve $\gamma : [0, 1] \to X$ is rectifiable of length L if $L = \sup \sum_{i=0}^{M-1} d_X(\gamma(t_i, t_{i+1})$ is finite, where the supremum is taken over all subdivisions $0 = t_0 < t_1 < \cdots < t_M = 1$ of the interval $[0, 1]$.

A group Γ acts *properly* on a metric space X if for all compact $K \subset X$, the set $\{g \in \Gamma \mid K \cap g \cdot K \neq \emptyset\}$ is finite.

Proposition 4.2. *Let (X, d_X) be a proper length space, and suppose that Γ acts properly and cocompactly on X. Then*

- *Γ is finitely generated;*
- *for any point $z \in X$, there is finite generating set $\mathcal{A}$ for Γ such that the map $Cay^1(\Gamma, \mathcal{A}) \to X$ defined on the vertices by $g \to g \cdot z$ is a quasi–isometry on the 0–skeleton.*

Remarks:

i) We shall see that the existence of a quasi–isometry between the Cayley graph and the space X is independent of the choice of finite generating set.

ii) The quasi–isometry can be extended over the edges by defining arbitrarily the images of the edges originating at the identity element, and then defining the images of the other edges using the Γ action (or alternatively by mapping each edge to the image of its initial vertex).

iii) Consider the standard hyperbolic structure on the surface S of genus two. That is, consider S as the quotient of the hyperbolic plane $\mathbb{H}^2$ by the action of a discrete group of hyperbolic isometries with fundamental domain a regular octagon with corner angle $\pi/4$. The plane $\mathbb{H}^2$ is a length space: in fact the surface S obtained in this way is a Riemannian manifold and between any pair of points there is a unique geodesic whose length is the distance between the points.

iv) Generalising the surface example, if M is a closed compact manifold where the distance between points is the infimum of the lengths of rectifiable curves between them (for instance when M has a Riemannian metric), then the universal covering space (with the lift of the distance from M) provides the length space X upon which $\pi_1(M)$ acts freely.

Proof. Let $C \subset X$ be a compact fundamental domain for the action of Γ, i.e. such that $\cup_{g\in\Gamma} g \cdot C = X$. Choose $z \in X$ and $r \geq 1$ such that $C \subset B_r(z)$.

We shall show that the finite set $\mathcal{A} = \{g \in \Gamma \mid \overline{B}_{3r}(z) \cap g \cdot \overline{B}_{3r}(z) \neq \emptyset\}$ generates Γ. Note that $y \in g \cdot \overline{B}_{3r}(z) \cap \overline{B}_{3r}(z) \implies g^{-1} \cdot y \in \overline{B}_{3r}(z) \cap g^{-1} \cdot \overline{B}_{3r}(z)$, so that $\mathcal{A}^{-1} = \mathcal{A}$.

To establish the quasi–isometry, we establish:

1) $\exists \lambda_1 > 0$ and $\epsilon_1 \geq 0$ s.t. $d_{\mathcal{A}}(1, g) \leq \lambda_1 d_X(z, g \cdot z) + \epsilon_1$.

2) $\exists \lambda_2$ such that $\forall g, g' \in \Gamma$, $d_X(g \cdot z, g' \cdot z) \leq \lambda_2 d_{\mathcal{A}}(g, g')$.

We begin by proving 1), and that $\mathcal{A}$ is a finite generating set.

As X is a length space, there is a rectifiable path $\gamma : [0, 1] \to X$ such that $\ell(\gamma) \leq d_X(z, g \cdot z) + 1$, and we can choose a subdivision $0 = t_0 < t_1 < \ldots t_M = 1$ such that $z = \gamma(0)$, $g \cdot z = \gamma(1)$, and $d_X(\gamma(t_i), \gamma(t_i + 1)) = r$ for $i < M - 1$ and $d_X(\gamma(t_{M-1}), \gamma(t_M)) \leq r$. Thus $Mr \leq \ell(\gamma) \leq (M + 1)r$, and $Mr \leq d_X(z, g \cdot z) + 1 \implies M \leq d_X(z, g \cdot z)/r + 1$.

Put $g_0 = 1$, $g_M = g$, and for each $1 \leq i \leq M - 1$, choose $g_i \in \Gamma$ such that $d_X(g_i \cdot z, \gamma(t_i)) \leq r$. This is possible as $\cup_g g \cdot B_r(z) = X$.

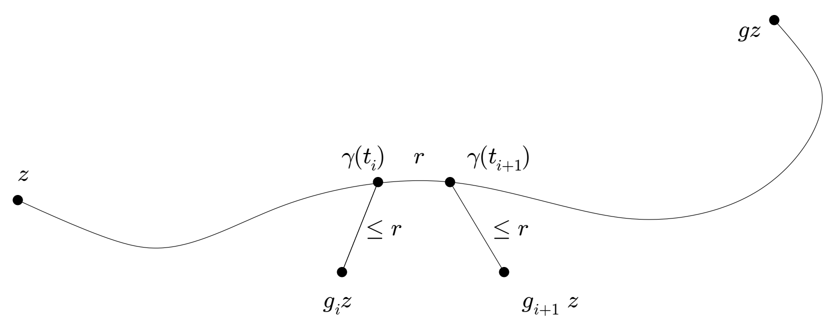


Figure 4.1: Proving finite generation

It follows that $d_X(g_i \cdot z, g_{i+1} \cdot z) \leq 3r \implies d_X(z, g_i^{-1} g_{i+1} \cdot z) \leq 3r$, and so $g_i^{-1} g_{i+1} \in \mathcal{A}$. But $g = g_M = 1.g_1.(g_1^{-1} g_2). \ldots .(g_{M-1}^{-1} g_M)$, and it follows that $\mathcal{A}$ generates Γ, and $d_{\mathcal{A}}(1, g) \leq M$. Moreover $d_X(z, g \cdot z) \geq Mr - 1 \geq d_{\mathcal{A}}(1, g)r - 1 \implies d_{\mathcal{A}}(1, g) \leq (1/r) d_X(z, g \cdot z) + (1/r)$.

To conclude, it suffices to take $\lambda_1 = \epsilon_1 = 1/r$.

We now procced to the proof of 2).
As $d_X(g \cdot z, g' \cdot z) = d_X(z, g^{-1} g' \cdot z)$, it suffices to consider the case $g' = 1$.

We know from 1) that $\mathcal{A}$ is a finite generating set. In terms of this generating set, there is a shortest word $a_1 a_2 \ldots a_n$ representing g in Γ; let $g_i = a_1 \ldots a_i$, for $i = 1, \ldots, n - 1$.

Then $d_X(z, g\cdot z) \le d_X(z, g_1\cdot z) + d_X(g_1\cdot z, g_2\cdot z) + \cdots + d_X(g_{n-1}\cdot z, g\cdot z) \le \lambda_2 n$ where $\lambda_2 = \max_{a\in\mathcal{A}} d_X(z, a\cdot z)$. But $n = d_{\mathcal{A}}(1, g)$, and 2) holds.

Summing up, we see from 1) that $d_{\mathcal{A}}(1, g) \le (1/r)d_X(z, g\cdot z) + (1/r)$, and from 2) that $1/\lambda_1 d_X(z, g\cdot z) \le d_{\mathcal{A}}(1, g)$. The inclusion of $Cay^1(\Gamma, \mathcal{A}) \to X$ is thus a $(\max\{1, \lambda_1, \lambda_2\}, r)$–quasi–isometry on the 0–skeleton.

□

The first (and perhaps the main) example of interest here concerns the maps obtained between Cayley graphs by changing the (finite) generating set of a finitely generated group. The above proof for good group actions apparently depends on the generating set used. The whole point of studying the geometry of the Cayley graph is to obtain properties of the group from the graph, so these properties must be invariant under change of finite generating set.

Proposition 4.3. *Let $\mathcal{A}$ and $\mathcal{B}$ be two finite generating sets for the group Γ.*

Then the Cayley graphs $Cay^1(\Gamma, \mathcal{A})$ and $Cay^1(\Gamma, \mathcal{B})$ are quasi–isometric via a map which is the identity on the vertices.

Proof. Let $\mathcal{A} = \{x_1, \ldots, x_p\}$, and $\mathcal{B} = \{y_1, \ldots, y_q\}$. For each x_i (respectively y_j), there is a word $u_i(\mathcal{B}) \in F(\mathcal{B})$ (resp. $v_j(\mathcal{A}) \in F(\mathcal{A})$) representing the same element as x_i (resp. y_j) in the group Γ.

Consider two elements $g, g'' \in \Gamma$, and suppose that $d_{\mathcal{A}}(g, g') = k$. Then there is a word $w(X) = x_{i_1}^{\epsilon_1} \ldots x_{i_k}^{\epsilon_k}$ in $F(\mathcal{A})$, such that $gw =_\Gamma g'$. Translating into the $\mathcal{B}$ generating set, gives the word $w(\mathcal{B}) = u_{i_1}^{\epsilon_1}(\mathcal{B}) \ldots u_{i_k}^{\epsilon_k}(\mathcal{B})$ such that $w(\mathcal{B}) =_\Gamma w(\mathcal{A})$, and the word $w(\mathcal{B})$ labels a path in the Cayley graph $Cay^1(\Gamma, \mathcal{B})$ from the vertex labelled g to the vertex labelled g'.

Let $K = \max\{\ell_{\mathcal{B}}(u_i(\mathcal{B})) \mid i = 1, \ldots, p\}$; then $\ell_{\mathcal{B}}(w(\mathcal{B})) \le Kk$ and so $d_{\mathcal{B}}(g, g') \le K d_{\mathcal{A}}(g, g')$.

In the same way, setting $K' = \max\{\ell_{\mathcal{A}}(v_j(\mathcal{A})) \mid j = 1, \ldots, q\}$, we see that $d_{\mathcal{A}}(g, g') \le K' d_{\mathcal{B}}(g, g')$.

Thus $(1/K')d_{\mathcal{A}}(g, g') \le d_{\mathcal{B}}(g, g') \le K d_{\mathcal{A}}(g, g')$.

In this way the identity map on the vertices is a quasi–isometry. To define a quasi–isometry on the edges, it suffices to extend to the identity map on the vertices to maps $\phi : Cay^1(\Gamma, \mathcal{A}) \to Cay^1(\Gamma, \mathcal{B})$ and $\psi : Cay^1(\Gamma, \mathcal{B}) \to Cay^1(\Gamma, \mathcal{A})$ taking the whole interior of each edge to either of its endpoints. Let t, t' be points in the interiors of two edges in $Cay^1(\Gamma, \mathcal{A})$, and consider $\phi(t), \phi(t')$ as elements of Γ, which can be viewed as a subset of $Cay^1(\Gamma, \mathcal{A})$ or of $Cay^1(\Gamma, \mathcal{B})$. Then $d_{\mathcal{A}}(t, t') \le 1$, so that $d_{\mathcal{B}}(\phi(t), \phi(t')) \le K d_{\mathcal{A}}(t, t') + 2K$.

Also clearly $d_{\mathcal{A}}(\psi\cdot\phi(t), t) \le 1$ and $d_{\mathcal{B}}(\phi\cdot\psi(s), s) \le 1$ for all $s \in Cay^1(\Gamma, \mathcal{B})$.

The maps ϕ and ψ are quasi–inverses, and the proof is complete. □

Definition 4.4. Two finitely generated groups are said to be quasi–isometric if they have quasi–isometric Cayley graphs with respect to some (and hence all by 4.3) finite generating sets. A group–theoretic property is said to be *geometric* if it is invariant under quasi–isometry of groups.

The first essential property is that of being finitely presented ([1]):

Proposition 4.5. *Let $\mathcal{A}$ and $\mathcal{B}$ be finite generating sets for the groups Γ and Γ'.*

If $Cay^1(\Gamma,\mathcal{A})$ and $Cay^1(\Gamma',\mathcal{B})$ are quasi–isometric and Γ is finitely presentable, then Γ' is also finitely presentable.

We shall in fact show something much stronger from which this proposition can be deduced: we shall show that quasi–isometric groups have comparable Dehn functions.

Definition 4.6. We say that two functions $f, g : \mathbb{N} \to \mathbb{R}$ are equivalent if there is a positive constant A such that for all $n \in \mathbb{N}$, $f(n) \leq Ag(An+A) + An + A$ and $g(n) \leq Af(An+A) + An + A$.

Notice that with this definition, linear functions are equivalent to constant functions (even to the zero function) and that all polynomials of degree $d > 1$ form an equivalence class, and all exponential functions form an equivalence class. It thus makes sense to talk about groups satisfying a linear, quadratic or exponential isoperimetric inequality, once we have shown the following result:

Theorem 4.7. [1] *Let $\mathcal{P} = \langle \mathcal{A} \mid R \rangle$ be a finite presentation of the group Γ.*

Let $\mathcal{B}$ be a finite generating set for the group Γ' such that $Cay^1(\Gamma,\mathcal{A})$ and $Cay^1(\Gamma',\mathcal{B})$ are quasi–isometric. Then there is a finite set of relators S for Γ' such that $\mathcal{Q} = \langle \mathcal{B} \mid S \rangle$ is a finite presentation for Γ', and the Dehn functions for the presentations $\mathcal{P}$ and $\mathcal{Q}$ are equivalent.

An immediate consequence of this is that "having solvable word problem" is a geometric property, and in particular:

Corollary 4.8. *If $\mathcal{P}$ and $\mathcal{Q}$ are finite presentations of the group Γ and the word problem is solvable for $\mathcal{P}$, then the word problem is solvable for $\mathcal{Q}$.*

Proof of the theorem. Let $\phi : Cay^1(\Gamma,\mathcal{A}) \to Cay^1(\Gamma',\mathcal{B})$ be a quasi–isometry and let $\psi : Cay^1(\Gamma',\mathcal{B}) \to Cay^1(\Gamma,\mathcal{A})$ be a quasi–inverse, so that for all vertices $g \in Cay^1(\Gamma',\mathcal{B})$, $d_{\mathcal{B}}(\phi \cdot \psi(g), g) \leq C$. Up to changing the quasi–isometry constants, we can suppose that vertices are sent to vertices, and that the vertices corresponding to the identity elements in each group are sent to each other.

Let $w = y_{j_1} \dots y_{j_k} \in F(\mathcal{B})$ be a word labelling a closed loop in $Cay^1(\Gamma',\mathcal{B})$ based at the vertex 1. For each initial segment $w_m = y_{j_1} \dots y_{j_m}$, for $m = 1, \dots, k$, let $v_m \in Cay^1(\Gamma',\mathcal{B})$ be the vertex represented by the word w_m. Consider the sequence of points $u_0 = 1, u_1 = \psi(v_1), \dots, u_k = \psi(v_k) = 1 \in Cay^1(\Gamma,\mathcal{A})$. As ψ is a (λ,ϵ)–quasi–isometry, and $d_{\mathcal{B}}(v_i, v_{i+1}) = 1$, we have $d_{\mathcal{A}}(u_i, u_{i+1}) \leq \lambda + \epsilon$.

There is therefore for each $i = 0, \dots, k$ a word $\alpha_i \in F(\mathcal{A})$ of length at most $\lambda + \epsilon$, such that there is a path in $Cay^1(\Gamma,\mathcal{A})$ labelled α_i from the vertex u_i to the vertex u_{i+1}. The product $w' = \alpha_0\alpha_1 \dots \alpha_k$ labels a loop in $Cay^1(\Gamma,\mathcal{A})$. There is therefore a van Kampen diagram D for w' over $\mathcal{P}$. The 1–skeleton of this diagram

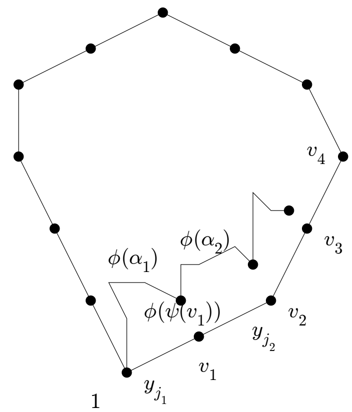


Figure 4.2: Part of the paths w and $\phi(\psi(w))$ in $Cay^1(\Gamma', \mathcal{B})$

maps into the Cayley graph $Cay^1(\Gamma, \mathcal{A})$. Applying $\phi : D^{(1)} \to Cay^1(\Gamma', \mathcal{B})$, corresponds to relabelling the diagram D, to get a diagram D' where each edge which was labelled by a generator in $\mathcal{A}$, is now labelled by a word in $F(\mathcal{B})$ of length at most $\lambda + \epsilon$. Each compact face of D was labelled by a word $r \in R$, and is now labelled by a word of length at most $\lambda\ell(r) + \epsilon$ in $F(\mathcal{B})$. The boundary of D, which was labelled by w' is now labelled by $w'' = \phi(\alpha_0)\phi(\alpha_1)\dots\phi(\alpha_k)$. Considering the path labelled w'' based at the vertex 1, the vertex reached by the initial segment $\phi(\alpha_0)\phi(\alpha_1)\dots\phi(\alpha_m)$ is the vertex $\phi \cdot \psi(v_m)$, and so is at distance at most C from the vertex v_m.

It follows that we can construct a van Kampen diagram for w over $\mathcal{P}'$ by adding to the diagram D' (the relabelled diagram D) for each $m = 0, \dots, k$ a diagram for the word $\phi(\alpha_m)y_{j_m}^{-1}h_m$, where h_m is a word of length at most C. In this way we obtain a van Kampen diagram for w of area $const.area_{\mathcal{P}}(\psi(w)) + const.\ell(w)$, and w is in the normal closure of set of relations in $F(\mathcal{B})$ of length at most $\max(C+1+\lambda+\epsilon, (\lambda+\epsilon)\max_{r\in R}\{\ell(r)\})$. As $\ell(\psi(w)) \leq const.\ell(w)$, this also shows an isoperimetric inequality for Γ gives an equivalent isoperimetric inequality for Γ'.

□

The special class of hyperbolic groups is the class of all finitely presented groups satisfying a linear isoperimetric inequality. An alternative definition is via the definition of “thin triangles”.

Definition 4.9. Let (X, d) be a geodesic metric space (length space) — i.e. where, between any two points there is a path (a “geodesic”) whose length is equal to the distance between the points.

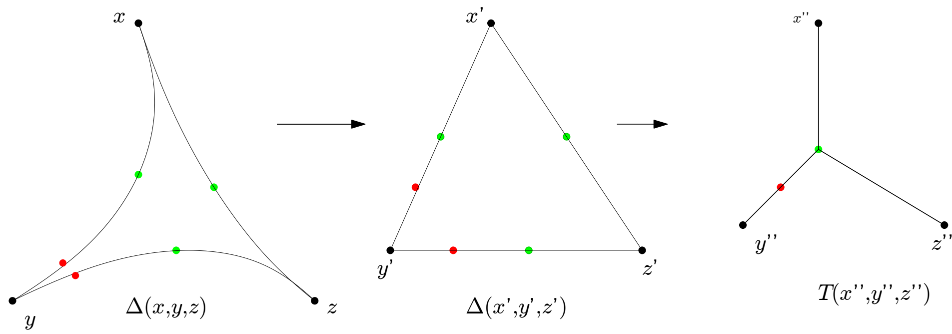
For any three points x, y, z there is a “geodesic triangle” $\Delta(x, y, z)$ formed by taking a geodesic between each pair of points. (There may be many such geodesics.)

Because of the triangle inequality, for each such triangle, there is a Euclidean triangle $\Delta(x', y', z')$ with the same side lengths. The Euclidean triangle maps onto a tripod $Y(x'', y'', z'')$, by collapsing the inscribed circle onto a point.

Let T_Δ be the composite map from the (edges of the) triangle $\Delta \to Y$. For a positive real number δ, we say that the triangle Δ is *δ–thin* if $\forall p \in Y$ the diameter $Diam(T^{-1}(p)) \leq \delta$.

The space X is said to be δ–hyperbolic if every geodesic triangle is δ–thin.

A finitely generated group Γ is said to be *hyperbolic* if it has a finite generating set $\mathcal{A}$ such that the Cayley graph $Cay^1(\Gamma, \mathcal{A})$ is δ–hyperbolic for some positive δ.



Sometimes δ–thin triangles are called uniformly δ–thin triangles (see [2]). An alternative definition is to say that a space is δ–hyperbolic space if for any geodesic triangle, each side lies in a δ–neighbourhood of the union of the other two sides. It is not to hard to show that the two definitions are equivalent (though it may be necessary to change the value of the constant δ). In order to show that all Cayley graphs (with respect to any finite generating sets) of a hyperbolic group are hyperbolic one must show that quasi–geodesic triangles (i.e. images of geodesic triangles under a quasi–isometry) are also thin, or some other equivalent result.

Before showing that hyperbolic groups satisfy a linear isoperimetric inequality, we shall first show that they satisfy a quadratic isoperimetric inequality, as this proof is simple and illustrates well the ideas of geometric group theory. A proof that they satisfy an isoperimetric inequality of type $n \log n$ is given in Section 6.2. We finally give in Proposition 4.11 a proof that they satisfy a linear isoperimetric inequality which is due to Noel Brady. It would suffice to show that geodesics in a hyperbolic group fellow travel, and then use Theorem 5.2.2 of Riley's notes to oobtain a linear isoperimetric inequality. In fact this same theorem states that a subquadratic isopermietric inequality implies a linear one (though this does use asymptotic cones!).

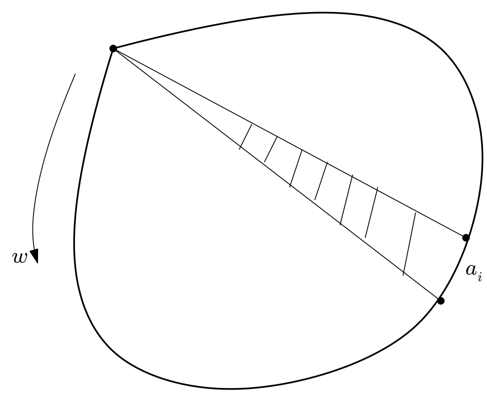
Proposition 4.10. *Let $\mathcal{A}$ be a finite generating set for the group Γ such that the Cayley graph $Cay^1(\Gamma, \mathcal{A})$ is δ–hyperbolic.*

Then Γ is finitely presentable and satisfies a quadratic isoperimetric inequality.

Proof. We shall suppose that no generator in $\mathcal{A}$ is trivial in Γ.

Let $w = a_1 \dots a_n \in F(\mathcal{A})$ be a word which represents the identity element of Γ. Then the word labels a closed loop based at the identity vertex of $Cay^1(\Gamma, \mathcal{A})$.

Let γ_i be a shortest word in $F(\mathcal{A})$ representing the element $a_1 \dots a_i$ in Γ. Then the paths based at 1 in $Cay^1(\Gamma, \mathcal{A})$ with labels γ_i and γ_{i+1} form a geodesic triangle, together with the edge labelled a_{i+1} based at the vertex $a_1 \dots a_i$ in $Cay^1(\Gamma, \mathcal{A})$. The fact that geodesic triangles are δ–thin means that this triangle can be decomposed as a collection of rectangles each of perimeter at most $2\delta + 2$ (the last is perhaps a triangle of perimeter at most $2\delta + 1$). There are at most $\max\{\ell(\gamma_i), \ell(\gamma_{i+1})\} \leq n/2$ of these rectangles.



It follows that the set of relations $R = \{r \in F(\mathcal{A}) \mid \ell(r) \leq 2\delta + 2,\ r =_\Gamma 1\}$ gives a finite presentation for Γ, and in terms of these relations, $area(w) \leq n^2/2$. $\square$

After this first simple proof, let us now show that in fact the group satisfies a linear isoperimetric inequality.

Proposition 4.11. *Let $\mathcal{A}$ be a finite generating set for the group Γ. If all geodesic triangles in $Cay^1(\Gamma, \mathcal{A})$ are δ–thin, then Γ is finitely presentable, has a Dehn presentation and satisfies a linear isoperimetric inequality*

Proof. The method used here (due to Noel Brady) is to show that "local geodesics" are like geodesics, and so Dehn's algorithm, with an appropriate set of relators solves the word problem for Γ.

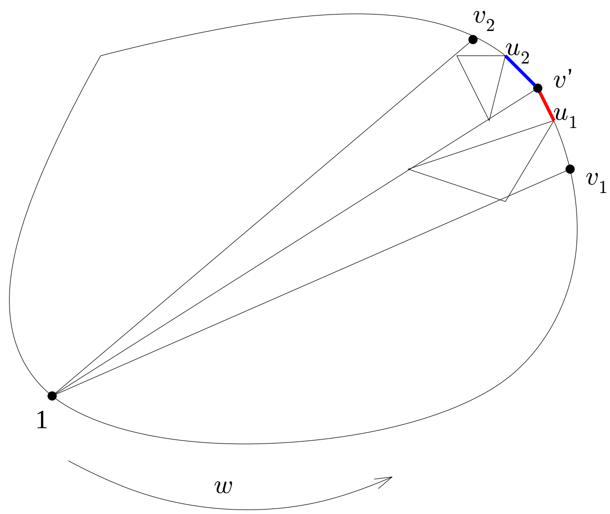
For $k > 0$, a word $w \in F(\mathcal{A})$ in $Cay^1(\Gamma, \mathcal{A})$ is a k *local geodesic* if all subwords of w of length at most k are geodesic. We shall show that $2\delta + 2$ local geodesics do not label loops:

If a word is not a $2\delta+2$ local geodesic, then it contains a subword v of length at most $2\delta+2$ such that there is a shorter word u such that $v =_\Gamma u$. Take as relators $R = \{r \in F(\mathcal{A}) \mid \ell(r) \leq 4\delta+3, r =_\Gamma 1\}$. Dehn's algorithm, using this set of relators, can thus be used to convert any word into a $2\delta+2$ local geodesic word representing the same element of the group (and in fact this can be done in time depending linearly on the length of the original word [2, 2.18] — it can even be done in *real time*, as has been shown by Holt and Röver [18]).

Claim: if $w =_\Gamma 1$ (and so w labels a loop in $Cay^1(\Gamma, \mathcal{A})$, then w is not a $2\delta+2$ local geodesic (i.e. it contains a subsegment of length at most $2\delta+2$ which is not geodesic).

Proof of the claim:

We argue by contradiction: let w be a non–empty word in $F(\mathcal{A})$ which represents the trivial element of Γ (labels a loop in $Cay^1(\Gamma, \mathcal{A})$) and suppose that w is a $2\delta+2$ local geodesic: the length of w is then at least $4\delta+4$.



Let γ be a loop in $Cay^1(\Gamma, \mathcal{A})$ based at the vertex 1 and labelled by the $2\delta+2$ local geodesic word w. Let v' be a vertex on γ furthest from the base point 1. This point is at distance at least $2\delta+1$ from 1, else w is trivial (the letter in the $2\delta+2$ position of w labels an edge ending at distance $2\delta+2$ from 1). Let v_1, v_2 be the vertices on γ before and after v' at distance $2\delta+1$ from v'.

Consider a geodesic triangle Δ_1 (resp. Δ_2) with vertices $1, v', v_1$ (resp. $1, v', v_2$) with one side the segment γ_1 (resp. γ_2) of γ between v' and v_1 (resp. v_2) of length $2\delta+1$. Let u_1 (resp. u_2) be the point on γ_1 (resp. γ_2) mapping to the central point of the tripod under the tripod map T_{Δ_1} (resp. T_{Δ_2}). Thus $|d_X(1, v_1) - d_X(1, v')| = |d_X(v_1, u_1) - d_X(u_1, v')|$.

If $d_X(u_1, v') < \delta+1$, then $d_X(v_1, u_1) \geq 2\delta+1-(\delta+1) \geq d_X(u_1, v')$ and so $d_X(v_1, 1) > d_X(v', 1)$ contradicting the choice of v'.

In the same way, we see that $d_X(u_2, v') \geq \delta + 1$. It follows that the points u'_1, u'_2 at distance $\delta + 1$ before and after v' on the segment of γ containing v' lie between u_1 and u_2. But u'_1 lies at distance $\delta + 1$ from v' on the geodesic from 1 to v', as does u'_2, and so $d_X(u'_1, u'_2) \leq 2\delta$, contradicting the fact that w is assumed to be a $2\delta + 1$ local geodesic.

$\square$

In this proof, if the path γ is not assumed to be a loop, what is proved is that the furthest point v' on γ from the initial point is within distance $2\delta + 1$ of the end of γ (i.e. the point u_2 cannot be constructed). Moreover, if one measures distance from any point $v \in Cay^1(\Gamma, \mathcal{A})$, rather than from the initial point of γ, one shows that the furthest point v' on γ from the point v lies within $2\delta + 1$ from one of the endpoints.

Gromov pointed out that a group satisfying a subquadratic isoperimetric inequality is in fact hyperbolic. Detailed proofs have been given by Papasoglu, Ol'shanskii and by Bowditch [4] (see also [6, p.422] for another presentation of Bowditch's proof). Another proof, using asymptotic cones, is given in Theorem 5.2.2 of Tim Riley's notes.

Chapter 5

Free nilpotent groups

The aim here is to give a lower bound for the isoperimetric inequality for free nilpotent groups, following [3]. The basic idea is to use a different method of estimating the area of a word in $\langle\langle R\rangle\rangle$. The method used has connections with group homology, but none of that theory is necessary in the constructions. Modulo a couple of elementary properties of nilpotent groups, complete proofs are given here.

When $\mathcal{P} = \langle \mathcal{A} \mid R\rangle$ is a finite presentation of the group Γ, and $Cay^1(\Gamma, \mathcal{A})$ is the Cayley graph, $\mathcal{R} = \langle\langle R\rangle\rangle$ can be identified with the fundamental group of $Cay^1(\Gamma, \mathcal{A})$. When Γ is not a finite group, this is an infinitely generated free group. We are interested in words $w \in F(\mathcal{A})$ such that $w \in \langle\langle R\rangle\rangle$, and thus there are conjugating elements p_i and relators $r_i \in R^{\pm 1}$ such that $w = \prod_{i=1}^{M} p_i r_i p_i^{-1}$. Recall that the area of w is the minimum such M.

Perhaps $\mathcal{R}$ is too complicated to be usable in computations. If we were simply to abelianise $\mathcal{R}$ and consider $\mathcal{R}/[\mathcal{R},\mathcal{R}]$, then we would still be dealing with an infinitely generated group. If, however we consider $\mathcal{R}/[\mathcal{R}, F]$, then we are considering a finitely generated abelian group. In this quotient, $r = prp^{-1}$ for all $r \in R$ and all $p \in F$, so the number of relators in $\mathcal{R}$ is an upper bound for the number of generators of $\mathcal{R}/[\mathcal{R}, F]$.

In the world of group homology, the exact sequence $1 \to \mathcal{R} \to F(\mathcal{A}) \to \Gamma \to 1$ leads to an exact sequence

$$\begin{array}{ccccccccc} 0 \to & \mathcal{R} \cap [F,F]/[\mathcal{R},F] & \to & \mathcal{R}/[\mathcal{R},F] & \to & F/[F,F] & \to & F/\mathcal{R}[F,F] & \to 0 \\ & \| & & \| & & \| & & \| & \\ 0 \to & H_2\Gamma & \to & \mathcal{R}/[\mathcal{R},F] & \to & H_1F & \to & H_1\Gamma & \to 0 \end{array}$$

and noting that H_1F is a free abelian group, we see that $\mathcal{R}/[\mathcal{R}, F] \cong H_2\Gamma \oplus \mathbb{Z}^k$ for some k. (In fact it is only the $H_2\Gamma$ part which is of interest to us, as is explained in [3]).

Now define a centralized isoperimetric inequality by measuring minimality in $\mathcal{R}/[\mathcal{R}, F]$. Define the centralized area of w to be $area^{cent}_{\mathcal{P}}(w) = \min\{M \mid w \in \prod_{i=1}^{M} p_i r^{\epsilon_i} p_i{}^{-1}[R, F]\}$. Thus we count just the number of times, with sign, that each relator occurs in a product of conjugates, ignoring the conjugating element involved.

Lemma 5.1. *Let $\mathcal{B} = \{y_1, \ldots, y_m\}$ be any finite set of generators for the abelian group $\mathcal{R}/[\mathcal{R}, F]$, and let $\ell_{\mathcal{B}}(w)$ be the minimal length of a word in these generators representing the element $w[\mathcal{R}, F]$ of $\mathcal{R}/[\mathcal{R}, F]$, and let $K = \max\{\ell_{\mathcal{B}}(r) \mid r \in R\}$.*

(1) Then $\ell_{\mathcal{B}}(w) \leq K.area^{cent}_{\mathcal{P}}(w) \leq K.area_{\mathcal{P}}(w)$;

(2) There is a positive constant C such that if $w[\mathcal{R}, F] = y^m[\mathcal{R}, F]$ and $y[\mathcal{R}, F]$ (and $w[\mathcal{R}, F]$) has infinite order in $\mathcal{R}/[\mathcal{R}, F]$ then $m \leq C.area^{cent}_{\mathcal{P}}(w)$.

Proof. (1) Write $w = \prod_{i=1}^{M} p_i r_i^{\epsilon_i} p_i{}^{-1}$ for some appropriate choices $p_i \in F$, $\epsilon_i = \pm 1$ and $r_i \in R$, with $M = area_{\mathcal{P}}(w)$. Removing the conjugating elements, we have $w[\mathcal{R}, F] = \prod_{i=1}^{M} r_i^{\epsilon_i}[\mathcal{R}, F]$. Also, if the centralized area is $area^{cent}_{\mathcal{P}}(w) = m$, then there are $q_j \in F$, $\beta_j = \pm 1$, and $s_j \in R$ such that $w[\mathcal{R}, F] = \prod_{j=1}^{m} q_j s_j^{\beta_j} q_j{}^{-1}[\mathcal{R}, F]$, and removing the conjugating elements we get $w[\mathcal{R}, F] = \prod_{j=1}^{m} s_j^{\beta_j}[\mathcal{R}, F]$. Then $\ell_{\mathcal{B}}(w) \leq \sum_j^m \ell_{\mathcal{B}}(s_j) \leq Km \leq KM$.

(2) As $\mathcal{R}/[\mathcal{R}, F]$ is a finitely generated abelian group, it is a direct sum of its torsion subgroup T and a free abelian group $\mathbb{Z}^k$. Choose a generating set $\mathcal{B}$ for $\mathcal{R}/[\mathcal{R}, F]$ consisting of a generating set for T and a basis for the $\mathbb{Z}^k$ summand. Mapping $\mathcal{R}/[\mathcal{R}, F]$ onto the $\mathbb{Z}^k$ summand, $w[\mathcal{R}, F] = y^m[\mathcal{R}, F]$ maps onto an m–th power, which is non–zero if $y[\mathcal{R}, F]$ has infinite order. Thus $m \leq \ell_{\mathcal{B}}(y^m) = \ell_{\mathcal{B}}(w) \leq K.area^{cent}_{\mathcal{P}}(w)$.

□

The point now is that in certain groups, it is possible to find words in F which are very short, but represent elements of $\mathcal{R}$ which are large powers in $\mathcal{R}/[\mathcal{R}, F]$. This is easy to do in nilpotent groups, as follows.

The lower central series of a group Γ is the sequence of groups $\Gamma_1 = \Gamma, \Gamma_2 = [\Gamma_1, \Gamma], \ldots, \Gamma_{k+1} = [\Gamma_k, \Gamma]$. A group is *nilpotent* if for some k, $\Gamma_k = 1$ (of class c if $\Gamma_c \neq 1$ and $\Gamma_{c+1} = 1$). Thus an abelian group is nilpotent of class 1. The simple k–fold commutators of Γ are those commutators of the form $[[\ldots[g_1, g_2], g_3], \ldots, g_k]$, which clearly lie in Γ_k. It is not hard to show by induction that if $X = \{x_1, \ldots, x_t\}$ is a set of generators for Γ, then the classes of the simple k–fold generators of the form $[[\ldots[\zeta_1, \zeta_2], \zeta_3], \ldots, \zeta_k]$ with $\zeta_j \in X$ generate Γ_k/Γ_{k+1} (see [21, 5.4]).

The free nilpotent group of class c on n generators is F/F_{c+1} where F is a free group on n generators.

Consider $\Gamma = F/\mathcal{R}$ with $\mathcal{R} = F_{c+1} = [F_c, F]$. Then $[\mathcal{R}, F] = F_{c+2}$, and $\mathcal{R}/[\mathcal{R}, F] = F_{c+1}/F_{c+2}$. According to the general result, this group is generated by the simple commutators.

We need the following basic facts about free nilpotent groups:

The c–fold simple commutator $[[\dots[[a,b],a],a],\dots,a]$ is a non–trivial element in F_c, and in F_c/F_{c+1} the commutator identities give
$[[\dots[[a^k,b^k],a^k],\dots],a^k] = [[\dots[[a,b],a],a],\dots,a]^{k^c} \mod F_{c+1}$.
For instance, the case of ordinary commutators:

$$\begin{aligned}[a^k,b] = a^k b a^{-k} b^{-1} &= a^{k-1} b a^{-(k-1)} b^{-1} (b a^{k-1} b^{-1} a b a^{-k} b^{-1}) \\ &= a^{k-1} b a^{-(k-1)} b^{-1} (b a^k b^{-1} a^{-1} (a b a^{-1} b^{-1}) a b a^{-k} b^{-1}) \\ &= [a^{k-1},b][a,b] \mod F_3\end{aligned}$$

By induction it follows that $[a^k,b^k] = [a,b^k]^k = [a,b]^{k^2} \mod F_3$. The general case is similar.

Thus, returning to our example of $\Gamma = F/\mathcal{R} = F/F_{c+1}$, we have the $c+1$–fold commutator $w_k = [[\dots[[a^k,b^k],a^k],\dots,a^k]$ is an element of F_{c+1} and in $F_{c+1}/F_{c+2} = \mathcal{R}/[\mathcal{R},F]$ this is a k^{c+1} power.

Thus the above lemmas on centralized area functions, $area^{cent}(w_k) \geq Ck^{c+1}$, while $\ell(w_k) \leq 2^{(c+1)}k$. But $area(w_k) \geq area^{cent}(w) \geq C'\ell(w_k)^{(c+1)}$ and so we have obtained a lower bound for the isoperimetric inequality which is polynomial of degree $c+1$ for the free nilpotent group of class c.

Chapter 6

Hyperbolic–by–free groups

As an example of how details of the structures of diagrams can help to give an interesting result, we look at N. Brady's result that there is a hyperbolic group containing a finitely presented non–hyperbolic subgroup. This example is a cyclic extension $1 \to K \to \Gamma \to \mathbb{Z} \to 1$. In this chapter we show that in examples of this type the kernel group K satisfies a polynomial isoperimetric inequality, following [15]. That is:

Theorem 6.1. *Let Γ be a split extension of a finitely presented group K by a finitely generated free group F, so one has the short exact sequence*

$$1 \to K \to \Gamma \to F \to 1.$$

If Γ is a hyperbolic group, then K satisfies a polynomial isoperimetric inequality.

The proof generalises easily to give an analgous result for groups satisfying a quadratic isoperimetric inequality — for details see [15]. The method of proof is to carefully study the form of van Kampen diagrams, using the area and intrinsic radius (see below) of a diagram over a presentation for Γ for a relation of K, viewed a relation of Γ, to give a diagram of bounded area over a presentation of K.

We need here the concept of *radius* of a diagram D, which is the maximum, over all vertices of D, of the number of edges in a shortest path in the 1–skeleton of D to the boundary δD. Properties of this function of diagrams are developed in section 5.2 of Tim Riley's notes. The important lemma we need is:

Lemma 6.2. *Let $\mathcal{P} = \langle \mathcal{A}; R \rangle$ be a finite presentation of a hyperbolic group Γ.*

Then there are constants $A, B > 0$ such that, for any relation $w \in F(\mathcal{A})$ with $\ell(w) \geq 1$, there is a van Kampen diagram over $\mathcal{P}$ of area at most $A\ell(w)(\log_2(\ell(w)) + 1)$ and of radius at most $B(\log_2(\ell(w)) + 1)$.

Proof. Consider a relation $w = c_1 \ldots c_M \in F(\mathcal{A})$ in Γ. Draw a circle in the plane, and subdivide into M vertices labelled by integers $i = 0, 1, \ldots, M-1$, which we

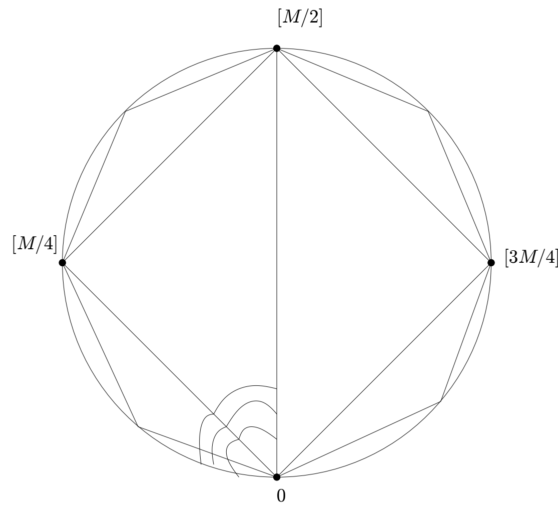


Figure 6.1: the first few subdivision triangles

consider as representatives for their equivalence classes mod n. Map this circle to a loop in the Cayley graph $Cay^1(\Gamma, \mathcal{A})$ based at the identity vertex via the word w. In the plane, join the vertices 0 and $[M/2]$ (the integer part of $M/2$) by a straight line, and extend the map to the Cayley graph over this arc by sending this arc to a geodesic γ_1 joining the appropriate vertices in $Cay^1(\Gamma, \mathcal{A})$.

For each integer $j = 2, \ldots, [\log_2(M) + 1]$, and for each $i = 1, \ldots, 2^j$, choose geodesics in $Cay^1(\Gamma, \mathcal{A})$, the *level* j geodesics, to label the straight lines joining the vertices $[(i-1)M/2^j]$ and $[iM/2^j]$ (some of these geodesics may degenerate to points for the last j). The level j triangles are then the geodesic triangles T_j^k, for $k = 1, \ldots 2^{j-1}$, with vertices $[2(i-1)M/2^j], [(2i-1)M/2^j]$ and $[2iM/2^j]$, and sides consisting of two level j geodesics $\gamma_j^{2i-1}, \gamma_j^{2i}$ and a level $j-1$ geodesic γ_{j-1}^i. At the final level take the edges in the loop w for the geodesics; at this level some of the triangles may degenerate. Notice that for each j, the sum of the lengths of the level j geodesics is at most M.

Suppose that K is δ–hyperbolic with respect to this presentation, so that each geodesic triangle can be decomposed into three triangles of area at most $\delta + 2$, a collection of rectangles of perimeter $2\delta + 2$, and a single central region of perimeter at most $3\delta + 3$. The number of these regions is at most half the perimeter of the triangle.

Filling in each of the level j triangles with these small triangles, rectangles and other central regions, construct a van Kampen diagram for the word w of area at most $A\ell(w)(\log_2(\ell(w)) + 1)$, where A is the maximum area of a minimal van Kampen diagrams over $\mathcal{P}$ for the relations of length at most $3\delta + 3$. If B' is the maximum radius of the minimal van Kampen diagrams over $\mathcal{P}$ for the relations

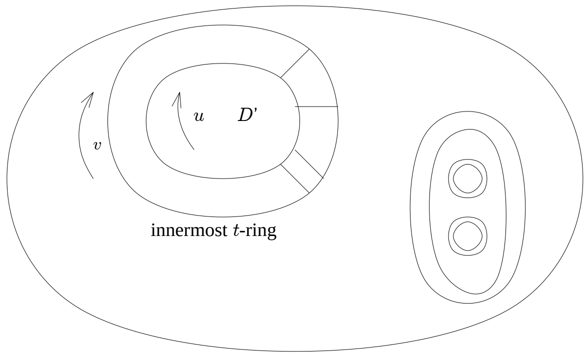


Figure 6.2: A diagram over $\mathcal{P}_\Gamma$ with some t–rings

of length at most $3\delta + 3$, then the radius of the constructed diagram is at most $\delta(\log_2(\ell(w)) + 1) + B' \leq B(\log_2(\ell(w)) + 1)$ for some B.

$\square$

In the same way, it is not hard to see how (see for instance Theorem 2.3.4 of Tim Riley's notes) the fellow traveller property for synchronously (respectively asynchronously) automatic groups gives constants $A, B > 0$ (resp. $C > 1,\ D > 0$) such that each relation w has a van Kampen diagrams of area at most $(A\ell(w)^2$ (resp. $C^\ell(w))$ and radius at most $B\ell(w)$ (resp. $D\ell(w)$).

Proof of the theorem. For simplicity we give the details for a cyclic extension. Fix $1 \to K \to \Gamma \to \mathbb{Z} \to 1$ a split extension, defined by the automorphism ϕ of K, and let $\mathcal{P}_K = \langle \mathcal{A} \mid R \rangle$ be a finite presentation for K, where $\mathcal{A} = \{a_1, \ldots, a_n\}$. It is clear that Γ has a presentation as an HNN extension with base group K and stable letter t, $\mathcal{P}_\Gamma = \langle a_1, \ldots, a_n, t \mid R, \{t^{-1}a_i t = w_\phi(a_i)\}\rangle$. In the general case, $\mathbb{Z}$ is replaced by a free group on k generators, and Γ is an HNN extension with k stable letters and k associated homomorphisms.

Let $\Phi : \mathcal{A}^* \to \mathcal{A}^*$ be the semigroup automorphism induced by ϕ restricted to $\mathcal{A}$. As ϕ is an automorphism, there is a semigroup homomorphism (acting as an inverse at the group level) $\Psi : \mathcal{A}^* \to \mathcal{A}^*$ such that $\Psi \cdot \Phi(a_i) =_K a_i$. For each of these, choose a van Kampen diagram $D_i, i = 1, \ldots, n$ over $\mathcal{P}_K$. To complete the proof of the main theorem, it remains to show how to obtain a diagram over $\mathcal{P}$ for a relation $w \in F(\mathcal{A})$ over $\mathcal{P}_\Gamma$ from a diagram over $\mathcal{P}_\Gamma$. As in the proof of Collins' Lemma (Lemma 2.10), the faces of the diagram over P_Γ corresponding to relations of the form $t^{-1}a_i t = w_\phi(a_i)$ combine to form annuli which we call t–rings, and fat arcs called t–corridors, meeting the boundary in t–edges.

Let $w \in F(\mathcal{A})$ be a relation over the presentation $\mathcal{P}_K$ for K. Then w is also a relation over the presentation $\mathcal{P}_\Gamma$ for Γ. Let D be a van Kampen diagram for w

over $\mathcal{P}_\Gamma$. As there are no occurrences of t in w, the faces of D coming from the relations of the form $t^{-1}a_it = \phi(a_i)$ form rings: there are no t–corridors.

Consider an innermost t–ring: i.e. inside the diagram D, there is a t–ring, i.e. an annulus A of adjacent faces all labelled by relators of the form $t^{-1}a_it = \phi(a_i)$ such that the component D' of the complement which does not meet ∂D (the inner component) contains no relators $t^{-1}a_it = \phi(a_i)$. Then D' is a diagram over $\mathcal{P}_K$ for the label u on the inner side of the annulus A. Let v be the label on the outer side of this annulus (the words u and v may be unreduced). There are now two cases to consider: either $v = \Phi(u)$ or $u = \Phi(v)$.

First note that applying the semigroup homomorphism Φ to the relator $r \in R$ gives a relator $\Phi(r)$. Let α be the maximum of the area of a mimimal diagram for $\Phi(r)$. In the same way there is a diagram of area at most β for each relation $\Psi(r)$. Claim: there is a van Kampen diagram for v over $\mathcal{P}_K$ of area $\leq \max\{\alpha, \beta\} area_{\mathcal{P}_K}(u)$

Case 1 : $v = \Phi(u)$. Subdivide each edge of the diagram D' for u, such that each edge which was originally labelled a_i is now labelled $\phi(a_i)$. Each face which was labelled $r_j \in R$ is now relabelled $\Phi(r_j)$, and each of these faces can be filled in by a diagram over $\mathcal{P}$ of area at most α.

Case 2 : $u = \Phi(v)$. Then in Γ, we have $v =_\Gamma \Psi(u)$. Subdivide and relabel as in case 1, but now each a_i–edge is relabelled $\Psi(a_i)$. Each face which was originally labelled r_j is now labelled $\Psi(r_j)$, and each of these can be filled in by a diagram of area at most β. This diagram for $\Psi(u)$ can be made into a diagram for v as follows. Noting that $u = \Phi(v)$, we have $\Psi(u) = \Psi{\cdot}\Phi(v)$ and that if $v = c_1 \ldots c_p$ with $c_j \in \mathcal{A}$, then $\Psi \cdot \Phi(v) = \Psi \cdot \Phi(c_1) \ldots \Psi \cdot \Phi(c_p)$, it suffices to add diagrams for each relation $c_j =_K \Psi \cdot \Phi(c_j)$ If γ is the maximum area of these diagrams, then there is a van Kampen diagram for v over $\mathcal{P}$ of area at most $\beta area_{\mathcal{P}}(u) + \gamma \ell(u) \leq \gamma' \ell(u)$.

To obtain a bound on the area of a $\mathcal{P}_K$ diagram for w it suffices to note that t–rings can be enclosed to a depth of at most the radius of D, and removing innermost t–rings one after the other multiplies area by at most $C = \max\{\alpha, \gamma'\}$.

Thus, as the original diagram D over $\mathcal{P}_\Gamma$ can be chosen of area at most $A\ell(w)(\log_2(\ell(w))+1)$, and of radius at most $B\log_2(\ell(w)+1)$, there is a $\mathcal{P}_K$ diagram for w of area at most $C^{B\log_2(\ell(w)+1)} A\ell(w)(\log_2(\ell(w)) + 1)$ which is bounded by a polynomial function of $\ell(w)$. $\square$

Bibliography

[1] J. Alonso, *Inégalités isoperimétriques et quasi-isométries*, C. R. Acad. Sci. Paris Série 1, 1990, **311**, 761-764.

[2] J. Alonso, T. Brady, D. Cooper, V. Ferlini, M. Lustig, M. Mihalik, M. Shapiro, and H. Short, *Notes on word hyperbolic groups*, in Group Theory from a Geometrical Viewpoint, ed. by E. Ghys, A. Haefliger, and A. Verjovsky, World Scientific, 3–63, 1991

[3] G. Baumslag, C. F. Miller, H. Short, *Isoperimetric inequalities and homology of groups*, Invent. Math. 113, (1993), 531–560.

[4] B. H. Bowditch, *A short proof that a subquadratic isoperimetric inequality implies a linear one*, Michigan Math. J. 42 (1995) 103-107.

[5] N. Brady, *Branched coverings of cubical complexes and subgroups of hyperbolic groups*, Jour. London Math. Soc., **60**, (1999), #2, 461–480.

[6] M. R. Bridson and A. Haefliger, *Metric spaces of non–positive curvature*, Springer–Verlag, 1999.

[7] A. Cayley, *On the Theory of Groups*, Proc. L.M.S., **9**, (1878), 126–133.

[8] I. Chiswell, D. Collins and J. Huebschmann, *Aspherical group presentations*. Math. Z. **178**, (1981), no. 1, 1-36.

[9] M. Coornaert, T. Delzant and A. Papadopoulos, *Notes sur les groupes hyperboliques de Gromov*, Lecture Notes in Mathematics, #1441 Springer–Verlag, 1990

[10] M. Dehn, *Über unendliche diskontinuierliche Gruppen*, Math. Ann, **71** (1912), 116–144; translated by J.Stillwell, *On infinite discontinuous groups*, in "Papers on Group Theory and Topology", Springer Verlag, 1987

[11] R. Fenn, *Techniques of Geometric Topology*, Cambridge Univ. Press, London Math. Soc. Lecture Notes No. 57, 1983

[12] R. Fenn and C. P. Rourke, *Klyachko's method and the solution of equations over torsion free groups*, l'Enseignement Mathématique 42 (1996), 49-74

[13] E. Ghys, *Les groupes hyperboliques*, Séminaire Bourbaki, 722, S.M.F., 1990

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

[15] S. M. Gersten and H. Short, *Some isoperimetric inequalities for kernels of free extensions*, Proceedings of the Congress in honour of John Stallings, Geométrica Dedicata, **92**, (2002), 63–72

[16] L. I. Greendlinger and M. D. Greendlinger, *On three of Lyndon's results about maps*, in Contributions to group theory, edited by K. I. Appel, P.E. Schupp and J. Ratcliffe, Contemp. Math., **33**, 212–213, A.M.S, 1984

[17] A. Hatcher, *Algebraic Topology*, C.U.P., 2002; also (legally) available at http://www.math.cornell.edu/~hatcher/AT/ATpage.html.

[18] D.R. Holt and K. Röver, *On real–time word problems*, J.L.M.S., **67**, (2003), 289-301

[19] A. Klyachko, *Funny property of sphere and equations over groups*, Comm. in Alg., **21** (7) (1993), 2555-2575

[20] R. Lyndon and P. E. Schupp, *Combinatorial group theory*, Springer Verlag, 1977

[21] W. Magnus, A. Karrass and D. Solitar, *Combinatorial Group Theory*, 3rd edition, Dover 1976

[22] C.F. Miller III and P.E. Schupp, *The geometry of HNN extensions*, Comm. Pure Appl. Math., **26**, (1973), 787–802

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

[24] W. D. Neumann and M. Shapiro, *Geometric group theory*, Notes for the ANU short course, 1996; Topology Atlas http://at.yorku.ca/i/a/a/i/13.htm

[25] Y. Ollivier, *Sharp phase transition theories for hyperbolicity of random groups*, Geom. Func Anal., **14**, (2004), 595–679

[26] Y. Ollivier, *On a small cancellation theorem of Gromov*, To appear in Bull. Belgian Math. Soc

[27] A Yu. Olshanskii, *Hyperbolicity of groups with subquadratic isoperimetric inequality*, Intern. J. of Algebra and Comput., (1991), 1(3): 281-289.

[28] A. S. Švarc, *Volume invariants of coverings*, Dokl. Akad. Nauk. SSSR, **105**, (1955), 32–34

[29] E. van Kampen, *On some lemmas in the theory of groups*, Amer. J. Math. **55**, (1933), 261–267

[30] C. M. Weinbaum, *The word and conjugacy problem for any tame prime alternating knot*, Proc. Amer. Math.Soc., **22**, (1971) 22–26