1

The geometry of the word problem

Martin R. Bridson

Introduction

The study of decision problems in group theory is a subject that does not impinge on most geometers' lives – for many it remains an apparently arcane region of mathematics near the borders of group theory and logic, echoing with talk of complexity and undecidability, devoid of the light of geometry. The study of minimal surfaces, on the other hand, is an immediately engaging field that combines the shimmering appeal of soap films with intriguing analytical problems; Plateau's problem has a particularly intuitive appeal. The first purpose of this article is to explain that despite this sharp contrast in emotions, the study of the large scale geometry of least-area discs in Riemannian manifolds is intimately connected with the study of the complexity of word problems in finitely presented groups.

Joseph Antoine Ferdinand Plateau was a Belgian physicist who, in 1873, published a stimulating account of his experiments with soap films [90]. The question of whether or not every rectifiable Jordan loop in 3-dimensional Euclidean space bounds a disc of minimal area subsequently became known as Plateau's Problem. This problem was solved by Jesse Douglas [37] and Tibor Radó [91] (independently) around 1930. In 1948 C.B. Morrey [75] extended the results of Douglas and Radó to a class of spaces that includes the universal covering of any closed, smooth Riemannian manifold M.

Once one knows that least-area discs exist in this generality, numerous questions come to mind concerning their local and global geometry (cf. [79], [86], and [71]). The questions on which we shall focus in this article concern the large-scale geometry of these discs: Can one bound the area of least-area discs in M by a function of the length of their boundaries? If so, what is the least such function? What happens to the asymptotic behaviour of this function when one perturbs the metric or varies M within its homotopy type? What can one say about the diameter of least-area discs? *etc.*

Remarkably, these questions turn out to be intimately connected with the nature of the word problem in the fundamental group of M, i.e. the problem of determining which words in the generators of the group equal the identity. The most important and striking connection of this type is given by the *Filling The-*

The author's research is supported by an EPSRC Advanced Fellowship

orem in Section 2: the smallest function $\mathrm{Fill}_0^M(l)$ bounding the area of least-area discs in terms of their boundary length has qualitatively the same asymptotic behaviour[1] as the *Dehn function* $\delta_{\pi_1 M}(l)$ of the fundamental group of M.

The Dehn function of a finitely presented group $\Gamma = \langle \mathcal{A} \mid \mathcal{R} \rangle$ measures the complexity of the word problem for Γ by giving the least upper bound on the number of defining relations $r \in \mathcal{R}$ that must be applied in order to show that a word w in the letters $\mathcal{A}^{\pm 1}$ is equal to $1 \in \Gamma$; the bound is given as a function of the length of w (see Paragraph 1.2).

The first purpose of this article is to give a thorough account of the Filling Theorem. The second purpose of this article is to sketch the current state of knowledge concerning Dehn functions. Thus, in Section 3, I shall explain what is known about the set of $\simeq$ classes of Dehn functions (equivalently, isoperimetric functions Fill_0^M of closed Riemannian manifolds), and I shall also describe what is known about the Dehn functions of various groups that are of geometric interest. In later sections we shall see a variety of methods for calculating Dehn functions (some geometric, some algebraic, and some purely combinatorial). Along the way we shall see examples of how the equivalence $\delta_{\pi_1 M} \simeq \mathrm{Fill}_0^M$ can inform in both directions (cf. (2.2) and Section 6).

Historical Background. The precise equivalence between filling functions of manifolds and complexity functions for word problems is a modern observation due to Mikhael Gromov, but this connection sits comfortably with the geometric origins of combinatorial group theory.

Topology and combinatorial group theory emerged from the same circle of ideas at the end of the nineteenth century. By 1910 Dehn had realized that the problems with which he was wrestling in his attempts to understand low-dimensional manifolds were instances of more general group-theoretic problems. In 1912 he published the celebrated paper in which he set forth the three basic decision problems that remained the main focus for combinatorial group theory throughout the twentieth century:

> *"The general discontinuous group is given by n generators and m relations between them.* [...] *Here there are above all three fundamental problems* [...]
> 1: [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.* [2: The Conjugacy Problem. 3: The Isomorphism Problem]
> *One is already led to them by necessity with work in topology. Each knotted space curve, in order to be completely understood, demands the solution of the three above problems in a special case.*[2]*"*

[1] More precisely, Fill_0^M is $\simeq$ equivalent to $\delta_{\pi_1 M}$ in the sense of 1.3.2.

[2] The special cases referred to here were not resolved fully until the early 1990s, and their ultimate solution rested on some of the deepest geometry and topology of the time, in particular the work of Thurston on the geometric nature of 3-manifolds.

In the present article I shall concentrate almost exclusively on the word problem, but in Section 8 I shall explain constructions that translate the complexity of word problems into conjugacy problems and isomorphism problems. These basic decision problems are all unsolvable in the absence of further hypotheses (see [72] for a survey of these matters) and in the spirit of Dehn's comments I should note that this undecidability has consequences for the study of manifolds. For example, the undecidability of the isomorphism problem for groups implies that there is no algorithm to recognise whether or not a closed 4-manifold (given by a finite triangulation, say) is homeomorphic to the 4-sphere [70].

Despite Dehn's early influence, the geometric vein in combinatorial group theory lacked prominence for much of the twentieth century (see [30] for a history up to 1980). A striking example of this neglect concerns a paper [61] written by E.R. van Kampen in 1931 which seems to have gone essentially unnoticed until rediscovered[3] by C. Weinbaum in the 1960s, just after Roger Lyndon [65] rediscovered the paper's main idea. This idea translates many questions concerning word problems into questions concerning the geometry of certain planar 2-complexes called *van Kampen diagrams* (see Section 4). This translation acts as a link between Riemannian filling problems and word problems. The work of Gromov [55], [56] gave full voice to the implications of this link. In the decade since Gromov's foundational work there has been a great deal of activity in this area and I hope that when the reader has finished the present article (s)he will have absorbed a sense of this activity and its achievements.

Contents. I have written this article with the intention that it should be accessible to graduate students and colleagues working in other areas of mathematics. It is organised as follows. In Section 1 we shall see how a naive head-on approach to the word problem leads to the definition of the Dehn function of a group. In Section 2 we introduce the 2-dimensional, genus-0 isoperimetric function of a closed Riemannian manifold M and state the theorem relating it to the Dehn function of $\pi_1 M$; the proof of this theorem is postponed until Section 5. This theorem is generally regarded as folklore – its validity has been assumed implicitly in many papers, but the absence of a detailed proof in the literature has been the source of comment and disquiet. The proof given here is self-contained. It is based on the notes from my lectures at the conferences in Durham, Lyon and Champoussin in the spring and summer of 1994. José Burillo and Jennifer Taback [26] have suggested an alternative proof, motivated by arguments in [42]. Both proofs rely on van Kampen's Lemma, which is proved in complete detail in Section 4.

Section 3 contains a brief survey describing the current state of knowledge about the nature of Dehn functions for groups in general as well as groups that are of particular geometric interest. We shall not prove the results in this section, but several of the key ideas involved are explained in subsequent sections.

[3] Van Kampen's article was next to the one in which he proved the Seifert-van Kampen Theorem.

Section 6 contains information about the classes of groups whose Dehn functions are linear or quadratic. We shall see that having a linear Dehn function is a manifestation of negative curvature. We shall also see that non-positive curvature is related to having a quadratic Dehn function, although the connection is much weaker than in the linear case.

The final section of this paper contains a brief discussion of different measures of complexity for the word problem, as well as constructions relating the word problem to the other basic decision problems of group theory.

There are three appendices to this paper. The first contains a description of some basic concepts in geometric group theory – this is included to make the arguments in the main body of the paper accessible to a wider audience. The second appendix describes some of the basic vocabulary of length spaces. The third appendix contains the proof of a technical result concerning the geometry of combinatorial discs; this result, which is original, is needed in Section 5.

Exercises are scattered throughout the text, some are routine verifications, some lead the diligent reader through proofs, and others are challenges intended to entice the reader along fruitful tangents.

This article is dedicated with deep affection to my tutor and friend Brian Steer. Between 1983 and 1986 Brian transformed me into a budding mathematician and thereby determined the course of my adult life.

1 The Word Problem

The purpose of this first section is to indicate why Dehn functions are fundamental to the understanding of discrete groups.

1.1 Presenting Groups that Arise in Nature

Suppose that one wishes to understand a group Γ that arises as a group of transformations of some mathematical object, for example isometries of a metric space. Typically, one might be interested in the group generated by certain basic transformations $\mathcal{A} = \{a_1, \ldots, a_n\}$. One then knows that arbitrary elements of Γ can be expressed as words in these generators and their inverses, but in order to gain a real understanding of the group one needs to know which pairs of words w, w' represent the same element of Γ, i.e. when $w^{-1}w' = 1$ in Γ. Words that represent the identity are called *relations*.

Let us suppose that the context in which our group arose is such that we can identify at least a few relations $\mathcal{R} = \{r_1, \ldots, r_m\}$. How might we use this list to deduce that other words represent the identity?

If a word w contains $r \in \mathcal{R}$ or its inverse as a subword, say[4] $w = w_1 r^{\pm 1} w_2$, then we can replace w by the shorter word $w' = w_1 w_2$, knowing that w' and w represent the same element of Γ. More generally, if r can be broken into (perhaps empty) subwords $r \equiv u_1 u_2 u_3$ and if $w \equiv w_1 u_2^{\pm 1} w_2$, then one knows that $w' \equiv w_1 (u_3 u_1)^{\mp 1} w_2$ equals w in Γ. Under these circumstances[5] one says **w' is obtained from w by applying the relator r.**

If we can reduce w to the empty word by applying a sequence of relators $r \in \mathcal{R}$, then we will have deduced that $w = 1$ in Γ. If such a sequence can be found for every word w that represents the identity – in other words, every relation in the group can be deduced from the set $\mathcal{R}$ – then the pair[6] $\langle \mathcal{A} \mid \mathcal{R} \rangle$ is called a *presentation* of Γ, and one writes[7] $\Gamma = \langle \mathcal{A} \mid \mathcal{R} \rangle$.

1.2 Attacking the Word Problem Head-On

A solution to the word problem in Γ is an algorithm that will decide which elements of the group represent the identity and which do not. If one can bound the number of relators that must be applied to a word w in order to show that $w = 1$, and this bound can be expressed as a computable function of the length of w, then one has an effective solution to the word problem. In order to quantify this idea precisely, one works with equalities in the free group $F(\mathcal{A})$.

Suppose that $w' = w_1 (u_3 u_1) w_2$ has been obtained from $w = w_1 u_2^{-1} w_2$ by applying the relator $r \equiv (u_1 u_2 u_3)^{-1}$. In Γ we have $w = w'$, while in the free

[4] We write $=$ for equality in the free group, and $\equiv$ when words are actually identical.

[5] At this point we are viewing words as elements of the free group $F(\mathcal{A})$, so implicitly we allow the insertion and deletion of subwords of the form aa^{-1}.

[6] If $\mathcal{R} = \{r_1, r_2, \ldots\}$, one often writes $\langle \mathcal{A} \mid r_1 = 1, r_2 = 1, \ldots \rangle$ instead of $\langle \mathcal{A} \mid \mathcal{R} \rangle$, particularly when this creates a desirable emphasis. Likewise, one may write $\langle \mathcal{A} \mid u_1 = v_1, u_2 = v_2, \ldots \rangle$, where $r_i \equiv u_i v_i^{-1}$.

[7] To assign a name to a presentation, P say, one writes $P \equiv \langle \mathcal{A} \mid \mathcal{R} \rangle$.

group $F(\mathcal{A})$ we have:

$$w \equiv w_1 u_2^{-1} w_2 \overset{\text{free}}{=} (x_1^{-1} r x_1)\ w_1 u_3 u_1 w_2 \equiv (x_1^{-1} r x_1)\ w',$$

where $x_1 := u_3^{-1} w_1^{-1}$. If w'' is a word obtained from w' by applying a further relator r', then there is an equality of the form $w \overset{\text{free}}{=} (x_1^{-1} r x_1)\ (x_2^{-1} r' x_2)\ w''$.

Proceeding in this manner, if we can reduce w to the empty word by applying a sequence of N relators from $\mathcal{R}$, then we will have an equality[8]

$$w \overset{\text{free}}{=} \prod_{i=1}^{N} x_i^{-1} r_i x_i, \tag{1.2.1}$$

where $r_i \in \mathcal{R}^{\pm 1}$ and $x_i \in F(\mathcal{A})$.

Thus we see that when one attacks the word problem head-on by simply applying a list of relators to a word w, one is implicitly expressing w as a product of conjugates of those relators. The ease with which one can expect to identify such an expression for w will vary according to the group under consideration, and in particular will depend very much on the number N of factors in a least such expression.

Definition 1.2.2 *Given a finite presentation $P \equiv \langle \mathcal{A} \mid \mathcal{R} \rangle$ defining a group Γ, we say that a word w in the letters $\mathcal{A}^{\pm 1}$ is* null-homotopic *if $w =_\Gamma 1$, i.e. w lies in the normal closure of $\mathcal{R}$ in the free group $F(\mathcal{A})$. We define the algebraic area of such a word to be*

$$\mathrm{Area}_a(w) := \min\{N \mid w \overset{\text{free}}{=} \prod_{i=1}^{N} x_i^{-1} r_i x_i \ \textit{with}\ x_i \in F(\mathcal{A}), r_i \in \mathcal{R}^{\pm 1}\}.$$

The Dehn function *of P is the function $\delta_P : \mathbb{N} \to \mathbb{N}$ defined by*

$$\delta_P(n) := \max\{\mathrm{Area}_a(w) \mid w =_\Gamma 1, |w| \leqslant n\},$$

where $|w|$ denotes the length of the word w.

1.3 The Dehn Function of a Group

Since we are really interested in groups rather than particular finite presentations of them, we would like to talk about the Dehn function of Γ rather than of P. The following exercise illustrates how the Dehn functions of different presentations of a group may vary.

Exercise 1.3.1 Show that the Dehn function of $\langle a \mid \emptyset \rangle$ is $\delta(n) \equiv 0$ and the Dehn function of $\langle a, b \mid b \rangle$ is $\delta(n) = n$. For each positive integer k find a presentation of $\mathbb{Z}$ with Dehn function $\delta(n) = kn$.

[8] This equality shows in particular that $\Gamma = \langle \mathcal{A} \mid \mathcal{R} \rangle$ iff the kernel of the natural map $F(\mathcal{A}) \to \Gamma$ is the normal closure of $\mathcal{R}$.

Definition 1.3.2 *Two functions $f, g : [0,\infty) \to [0,\infty)$ are said to be $\simeq$ equivalent if $f \preceq g$ and $g \preceq f$, where $f \preceq g$ means that there exists a constant $C > 0$ such that $f(l) \leqslant C\, g(Cl + C) + Cl + C$ for all $l \geqslant 0$.*

One extends this equivalence relation to functions $\mathbb{N} \to [0,\infty)$ by assuming them to be constant on each interval $[n, n+1)$.

The relation $\simeq$ preserves the asymptotic nature of a function. For example, if $p > 1$ then $n^p \not\simeq n^p \log n$, and $n^p \simeq n^q$ implies $q = p$; likewise, $n^p \not\simeq 2^n$ and $2^{2^n} \not\simeq 2^n$. But $\simeq$ identifies all polynomials of the same degree, and likewise all single exponentials ($k^n \simeq K^n$ for all constants $k, K > 1$).

Proposition 1.3.3 *If the groups defined by two finite presentations are isomorphic, the Dehn functions of those presentations are $\simeq$ equivalent.*

Proof First we consider what happens when we add redundant relators $\mathcal{R}'$ to a finite presentation $P \equiv \langle \mathcal{A} \mid \mathcal{R} \rangle$. Let $P' \equiv \langle \mathcal{A} \mid \mathcal{R} \cup \mathcal{R}' \rangle$. To say that the new relators $r \in \mathcal{R}'$ are redundant means that each can be expressed in the free group $F(\mathcal{A})$ as a product Π_r of (say m_r) conjugates of the old relators $\mathcal{R}^{\pm 1}$. Let m be the maximum of the m_r.

If a word $w \in F(\mathcal{A})$ is a product of N conjugates of relators from $\mathcal{R} \cup \mathcal{R}'$ and their inverses, then by substituting Π_r for each occurrence of $r \in \mathcal{R}'$ in this product we can rewrite w (freely) as a product of at most mN conjugates of the relators $\mathcal{R}^{\pm 1}$. Since it is obvious that the area of w with respect to P' is not greater than its area with respect to P, we have $\delta_{P'}(n) \leqslant \delta_P(n) \leqslant m\, \delta_{P'}(n)$ for all $n \in \mathbb{N}$. Hence $\delta_P \simeq \delta_{P'}$.

Next we consider what happens when we add finitely many generators and relators to P. Suppose that we add generators $\mathcal{B}$, and add one relator bu_b^{-1} for each $b \in \mathcal{B}$, where u_b is a word in $F(\mathcal{A})$ that equals b in the group being presented. Let P'' be the resulting presentation. Let M be the maximum of the lengths of the words u_b.

Given a null-homotopic word $w \in F(\mathcal{A} \cup \mathcal{B})$, we first apply the new relators to replace each occurrence of each letter $b \in \mathcal{B}$ with the word u_b. The result is a word in $F(\mathcal{A})$ that has length at most $M|w|$, and this word may be reduced to the empty word by applying at most $\delta_P(M|w|)$ relators from $\mathcal{R}$. Thus $\delta_{P''} \preceq \delta_P$.

We claim that $\delta_P(n) \leqslant \delta_{P''}(n)$ for all $n \in \mathbb{N}$. To prove this claim we must show that if a word $w \in F(\mathcal{A})$ can be expressed in $F(\mathcal{A} \cup \mathcal{B})$ as a product Π of at most N conjugates of the given relators, then it can also be expressed in $F(\mathcal{A})$ as a product of at most N conjugates of the relators $\mathcal{R}^{\pm 1}$. To see that this is the case, one simply looks at the image of Π under the retraction $F(\mathcal{A} \cup \mathcal{B}) \to F(\mathcal{A})$ that sends each $b \in \mathcal{B}$ to u_b.

In general, given two finite presentations $P_1 \equiv \langle \mathcal{A} \mid \mathcal{R} \rangle$ and $P_2 \equiv \langle \mathcal{B} \mid \mathcal{R}' \rangle$ of a group G, one considers the presentation of G that has generators $\mathcal{A} \cup \mathcal{B}$ and relators $\mathcal{R}, \mathcal{R}', \{bu_b^{-1} \mid b \in \mathcal{B}\}$ and $\{av_a^{-1} \mid a \in \mathcal{A}\}$, where u_b (respectively v_a) is a word in $F(\mathcal{A})$ (respectively $F(\mathcal{B})$) that equals b (respectively a) in G. The first two steps of the proof imply that the Dehn function of this presentation is equivalent to that of both P_1 and P_2. □

The first detailed proof of (1.3.3) in the literature is due to Steve Gersten [45]. A more general result given in Appendix B (Proposition A.1.7) lends a geometric perspective to the equivalence in (1.3.3).

Isoperimetric Inequalities and δ_Γ. In the light of the preceding proposition we may talk of *"the" Dehn function* of a finitely presented group Γ, denoted δ_Γ, with the understanding that this is only defined up to $\simeq$ equivalence.

One says that Γ satisfies a *quadratic isoperimetric inequality* if $\delta_\Gamma(n) \preceq n^2$. Linear (also polynomial, exponential, *etc.*) isoperimetric inequalities are defined similarly.

A finitely generated group is said to have a solvable word problem if there is an algorithm that decides which words in the generators represent the identity and which do not. Readers who are familiar with the rudiments of decidability should treat the following statement as an exercise, and those who are not may treat it as a definition.

Proposition 1.3.4 *A finitely presentable group Γ has a solvable word problem if and only if the Dehn function of every finite presentation of Γ is computable (i.e. is a recursive function).*

Exercise 1.3.5 Two groups are said to be commensurable if they have isomorphic subgroups of finite index. Deduce from the Filling Theorem (Section 2) that the Dehn functions of commensurable finitely-presented groups are $\simeq$ equivalent. (Hint: Use covering spaces.)

The reader might find it instructive to investigate how awkward it is to prove this fact algebraically.

2 The Isoperimetric Function Fill_0^M of a Manifold

Let M be a closed, smooth, Riemannian manifold. In this section we shall describe the filling function Fill_0^M and its relationship to the Dehn function of the fundamental group of M.

2.1 The Filling Theorem

Let D be a 2-dimensional disc and let S^1 be its boundary circle. Let M be a smooth, complete, Riemannian manifold. Let $c : S^1 \to M$ be a null-homotopic, rectifiable loop and define $\mathrm{FArea}(c)$ to be the infimum of the areas[9] of all Lipschitz maps $g : D \to X$ such that $g|_{\partial D}$ is a reparameterization[10] of c. If this

[9] The situations that we shall be considering are sufficiently regular as to render all standard notions of area equivalent; for definiteness one could take 2-dimensional Hausdorff measure, or the notion of (Lebesgue) area in spaces with upper curvature bounds introduced by Alexandrov [1] and refined by Nikolaev (see [11] and [22] page 425).

[10] When working with filling problems it is usually better to consider loops that are equivalent in the sense of Frechet, but this technicality will have no bearing here.

infimum is attained by a (not necessarily injective) map $f : D \to M$ then, blurring the question of reparameterization, we say that f is a *least-area filling* of the loop $c = f|_{\partial D}$, or simply that f is a *least-area disc.*

If M is the universal covering of a closed manifold, then the existence of least-area discs (for embedded loops) is guaranteed by Morrey's solution to Plateau's problem [75].

Definition 2.1.1 *Let M be a smooth, complete, Riemannian manifold. The* genus zero, 2-dimensional, isoperimetric function *of M is the function* $[0, \infty) \to [0, \infty)$ *defined by*

$$\mathrm{Fill}_0^M(l) := \sup\{\mathrm{FArea}(c) \mid c : S^1 \to M \textit{ null-homotopic, length}(c) \leqslant l\}.$$

One of the main purposes of this article is to provide a detailed proof of the following fundamental equivalence:

2.1.2 Filling Theorem. *The genus zero, 2-dimensional isoperimetric function* Fill_0^M *of any smooth, closed, Riemannian manifold M is $\simeq$ equivalent to the Dehn function $\delta_{\pi_1 M}$ of the fundamental group of M.*

Remark 2.1.3 A similar statement holds with regard to isoperimetric functions of more general classes of spaces with upper curvature bounds (in the sense of Alexandrov [22]) but we shall not dwell on this point as we do not wish to obscure the main ideas with the technicalities required to set-up the required definitions. Nevertheless, in our proof of the filling theorem we shall make a point of isolating the key hypotheses so as to render these generalisations straightforward (cf. 5.2.2). In particular we avoid using any facts concerning the regularity of solutions to Plateau's problem in the Riemannian setting.

We postpone the proof of the Filling Theorem to Section 5, but we take a moment now to remove a concern about the definition of Fill_0^M: *a priori* the supremum in the definition of $\mathrm{Fill}_0^M(l)$ could be infinite for certain values of l even if M is compact, but in fact it is not.

Lemma 2.1.4 *If M is compact, the* sup *in the definition of* $\mathrm{Fill}_0^M(l)$ *is finite for all $l \geqslant 0$.*

Proof If the sectional curvature of M is bounded above by $k > 0$ then any null-homotopic loop in M of length $l < 2\pi/\sqrt{k}$ bounds a disc whose area is at most the area $A(k, l)$ of the disc enclosed by a circle of length l on the sphere of constant curvature k. Indeed Reshetnyak [93] proved that this bound holds in any complete geodesic space of curvature $\leqslant k$ (cf. appendix to [71]).

Let $\rho > 0$ be less than the injectivity radius of M, fix a finite set S so that every point of M lies in the $\rho/3$-neighbourhood of S and let $e_{x,x'} : [0, 1] \to M$ be the constant speed geodesic joining each $x, x' \in S$ with $d(x, x') < \rho$.

Given any constant-speed loop $c : [0, 1] \to M$, one can associate to it the concatenation $\hat{c} = e_{x_0,x_1} \cdots e_{x_n,x_0}$ where n is the least integer greater than $3l(c)/\rho$ and $x_i \in S$ is such that $d(x_i, c(i/n)) < \rho/3$ (cf. figure 5.1.2).

By construction, $|\mathrm{FArea}(c) - \mathrm{FArea}(\hat{c})| \leqslant n\,A(k, 2\rho)$ and $l(\hat{c}) \leqslant 3l(c) + \rho$. It follows that the $\simeq$ class Fill_0^M remains unchanged if instead of quantifying over all rectifiable loops c one quantifies only over loops that are concatenations of the loops $e_{x,x'}$. For all $L > 0$, there are only finitely many such edge-loops of length $\leqslant L$, so in particular $\mathrm{Fill}_0^M(l)$ is finite for all l. □

Remark 2.1.5 The reduction to piecewise-geodesic loops in the above proof exemplifies the fact that if one is concerned only with the $\simeq$ class of Fill_0^M then there is no harm in restricting one's attention to well-behaved sub-classes of rectifiable loops.

2.2 Filling in Heisenberg Groups

The results described in this paragraph are due to Mikhael Gromov. We present them here in order to give an immediate illustration of how one can exploit the equivalence $\mathrm{Fill}_0^M \simeq \delta_{\pi_1 M}$.

Let $n = 2m + 1$. The n-dimensional Heisenberg group $\mathcal{H}_n$ is the group of $(m+1)$-by-$(m+1)$ real matrices of the form:

$$\begin{pmatrix} 1 & x_1 & \dots & x_{m-1} & z \\ 0 & 1 & 0 & 0 & y_1 \\ \vdots & & & \vdots & \vdots \\ 0 & 0 & \dots & 1 & y_{m-1} \\ 0 & 0 & \dots & 0 & 1 \end{pmatrix}.$$

$\mathcal{H}_n$ is a nilpotent Lie group. Its Lie algebra L is generated by $X_1, \dots, X_{m-1}$, $Y_1, \dots, Y_{m-1}, Z = X_m = Y_m$ with relations $[X_i, Y_j] = [X_i, X_j] = [Y_i, Y_j] = 0$ for all $i \neq j$ and $[X_i, Y_i] = Z$ for $i = 1, \dots, m-1$. There is a natural grading $L = L_1 \oplus L_2$, where L_2 is spanned by Z and L_1 is spanned by the remaining X_i and Y_i.

The translates of L_1 by the left action of $\mathcal{H}_n$ form a sub-bundle T_1 of the tangent bundle of $\mathcal{H}_n$. (This codimension-1 sub-bundle gives the standard contact structure on $\mathcal{H}_n$.) A curve or surface mapped to $\mathcal{H}_n$ is said to be *horizontal* if it is differentiable almost everywhere and its tangent vectors lie in T_1. Every smooth curve c in $\mathcal{H}_n$ can be approximated by a horizontal curve whose length is arbitrarily close to that of c. The question of whether every horizontal loop bounds a horizontal disc ("the horizontal filling problem") is delicate, and it is here that we find a connection with Dehn functions.

The following result is an application of the theory developed by Gromov in Section 2.3.8 of his book on partial differential relations [54] and is explained on page 85 of [56].

Proposition 2.2.1 *If every horizontal loop in $\mathcal{H}_n$ can be filled with a horizontal disc, then $\mathrm{Fill}_0^{\mathcal{H}_n}(l) \simeq l^2$.*

The idea of the proof is as follows. First one must argue that there is a constant C such that any curve of length $\leqslant 1$ can be filled with a horizontal disc of area at most C. Then one considers the 1-parameter family of maps $h_t = \exp \circ \lambda_t \circ \exp^{-1} : \mathcal{H}_n \to \mathcal{H}_n$, where the Lie-algebra homomorphism $\lambda_t : L \to L$ is multiplication by $t \in [0,1]$ on L_1 and by t^2 on L_2. Note that h_t multiplies the length of horizontal curves by t and the area of horizontal discs by t^2.

Given a horizontal loop $c : S^1 \to \mathcal{H}_n$ of length $l > 1$, we consider $h_{1/l} \circ c$. One can fill this horizontal loop of length 1 with a horizontal disc $f_0 : D \to H_n$ of area at most C and hence obtain a horizontal disc $f := h_{1/l}^{-1} \circ f_0$ of area $\leqslant Cl^2$ that fills c. Since arbitrary loops can be approximated by horizontal loops, it follows that $\mathcal{H}_n$ satisfies a quadratic isoperimetric inequality.

The *integer Heisenberg group* H_n consists of those matrices in $\mathcal{H}_n$ that have integer entries. The subgroup $H_n \subset \mathcal{H}_n$ is discrete, torsion-free and cocompact, hence $M := H_n \backslash \mathcal{H}_n$ is a compact Riemannian manifold with universal covering $\mathcal{H}_n$, and $\delta_{H_n} \simeq \mathrm{Fill}_0^M = \mathrm{Fill}_0^{\mathcal{H}_n}$.

Gromov proves that the horizontal filling problem is solvable in $\mathcal{H}_n$ if and only if $n \geqslant 5$. It therefore follows[11] from the Filling Theorem and the above proposition that the integral Heisenberg group H_n has a quadratic Dehn function if $n \geqslant 5$. On the other hand, it is not hard to show by various combinatorial means (see 3.1.4 and 3.3.1 below) that the Dehn function of H_3 is cubic, so from the Filling Theorem and the above proposition one gets a proof of the easier "only if" implication in Gromov's theorem: $\mathcal{H}_3$ contains horizontal loops of finite length that cannot be filled with a horizontal disc.

3 Which Functions are Dehn Functions?

The most fundamental question concerning isoperimetric inequalities for finitely presented groups is that of determining which $\simeq$ equivalence classes of functions arise as Dehn functions. The struggle to solve this question was a major theme in geometric group theory in the 1990s. In this section I shall explain why this struggle is almost over. I shall also describe what is known about the Dehn functions of certain groups that are of special interest in geometry and topology.

Section 7 contains a sample of the techniques that were developed to establish the results quoted in the present section.

3.1 The Isoperimetric Spectrum

The development of knowledge concerning the nature of Dehn functions is best explained in terms of how the set of numbers

$$\mathrm{IP} = \{\rho \in [1,\infty) \mid f(n) = n^\rho \text{ is } \simeq \text{ a Dehn function}\}$$

came to be understood. This set is called the *isoperimetric spectrum*.

Since there are only countably many finite presentations of groups, Proposition 1.3.3 implies that there are only countably many $\simeq$ classes of Dehn func-

[11] For a self-contained proof along these lines see Allcock [2]. More recently, a purely combinatorial proof has been discovered by Ol'shanskii and Sapir [83].

tions. Thus, intriguingly, IP is a naturally arising countable set of positive numbers.

Integer Exponents. In Section 6 we shall discuss the class of groups that have linear Dehn functions. The following exercises describe the simplest examples from this class.

Exercises 3.1.1 (i) Finite groups and free groups have linear Dehn functions.

(ii) Let $\mathbb{H}^2$ denote the hyperbolic plane. There is a constant $C > 0$ such that for all $l > 1$, each loop in $\mathbb{H}^2$ of length $\leqslant l$ bounds a disc of area $\leqslant Cl$.

(iii) Every finitely generated group that acts properly by isometries on $\mathbb{H}^2$ has a linear Dehn function. (Hint: If the action is cocompact you can use (ii). If the action is not cocompact, argue that the group must have a free subgroup of finite index.)

In Section 6 we shall also describe what is known about the class of groups that have quadratic Dehn functions. Finitely generated abelian groups provide the easiest examples in this class.

Example 3.1.2 The Dehn function of $P \equiv \langle a, b \mid [a, b] \rangle$ is quadratic. More precisely, $(l^2 - 2l - 3) \leqslant 16\, \delta_P(l) \leqslant l^2$, the upper bound being attained in the case of words of the form $a^{-n}b^{-n}a^n b^n$.

Exercise 3.1.3 Prove that the inequality in (3.1.2) holds for the natural presentation of any free abelian group $\mathbb{Z}^r$, $r \geqslant 2$, and that it is optimal. (Hint: Given a word w that equals the identity in $\mathbb{Z}^r$, focus on a specific generator a and move all occurrences of $a^{\pm 1}$ to the left in w by applying the relators $[a, b] = 1$, freely reducing the resulting word whenever possible. Repeat for each generator and count the total number of relators applied — cf. Paragraph 1.2. If you have trouble with the lower bound, look at Section 7.)

In about 1988 Bill Thurston [42] and Steve Gersten [45] proved that the 3-dimensional Heisenberg group H_3 has a cubic Dehn function (see paragraph 2.2 and Theorem 3.3.1).

It now seems odd to report that there was a lull of a few years before people discovered sequences of groups $(\Gamma_d)_{d\in\mathbb{N}}$ such that the Dehn function of Γ_d is polynomial of degree d. Such sequences were described by a number authors at about the same time – Gromov [56], Baumslag, Miller and Short [10], and Bridson-Pittet [23]. The following result, proved by Bridson and Gersten in [21], provides many such sequences, and the literature now contains examples with all manner of additional properties (e.g. having Eilenberg-Maclane spaces of specified dimension [16]).

Theorem 3.1.4 *The Dehn function of each semi-direct product of the form* $\mathbb{Z}^n \rtimes_\phi \mathbb{Z}$ *is* $\simeq$ *either a polynomial or an exponential function. It is polynomial if and only if all of the eigenvalues of* $\phi \in \mathrm{GL}(n, \mathbb{Z})$ *are roots of unity, in which* *case the degree of the polynomial is* $c+1$*, where* c *is the size of the largest elementary block in the Jordan form of* ϕ*.*

Notice that groups of the form $\mathbb{Z}^n \rtimes_\phi \mathbb{Z}$ are precisely those that arise as fundamental groups of torus bundles over the circle, and hence the above theorem classifies the isoperimetric functions Fill_0^M of such bundles.

The appearance of the Jordan form in the above theorem is connected to the following facts (cf. 7.1.4).

Exercises 3.1.5 (i) If a matrix $\phi \in \mathrm{GL}(n,\mathbb{Z})$ does not have an eigenvalue of absolute value greater than 1, then all of its eigenvalues are N-th roots of unity, where N depends only on n. (Hint, [21], page 7: Let $P \subset \mathbb{Z}[x]$ be the set of monic polynomials of degree n whose roots all lie on the unit circle. P is finite. If the characteristic polynomial of ϕ lies in P then so does that of each power ϕ^r.)

(ii) Regard $\mathrm{GL}(n,\mathbb{Z})$ as a subset of $\mathbb{R}^{n^2}$ and fix a norm on $\mathbb{R}^{n^2}$. Prove that $m \mapsto \|\phi^m\|$ is $\simeq$ equivalent to an exponential function or a polynomial of degree $c-1$, where c is the size of the largest elementary block in the Jordan form of ϕ.

Filling the Gaps in IP**.** The following theorem is due to Gromov [55]. Detailed proofs were given by Ol'shanskii [81], Bowditch [14] (also [22] page 422) and Papasoglu [87].

Theorem 3.1.6 *If the Dehn function of a group is sub-quadratic (i.e.* $\delta_\Gamma(n) = o(n^2)$*) then it is linear (*$\delta_\Gamma(n) \simeq n$*). Thus* $\mathrm{IP} \cap (1,2)$ *is empty.*

This theorem begs the question of what other gaps there may be in the isoperimetric spectrum, or indeed whether there are any non-integral isoperimetric exponents at all. This last question was settled by the discovery of the *abc* groups [19]. These groups are obtained by taking three torus bundles over the circle (each of a different dimension) and amalgamating their fundamental groups along central cyclic subgroups.

The basic building block is $G_c = \mathbb{Z}^c \rtimes_{\phi_c} \mathbb{Z}$, where $\phi_c \in \mathrm{GL}(c,\mathbb{Z})$ is the unipotent matrix with ones on the diagonal and super-diagonal and zeros elsewhere. G_c has presentation:

$$\langle x_1,\dots,x_c,t \mid [x_i,x_j]=1 \text{ for all } i,j,\ [x_c,t]=1,\ [x_i,t]=x_{i+1} \text{ if } i<c\rangle. \quad (3.1.7)$$

Notice that the centre of G_c is the infinite cyclic subgroup generated by x_c. To emphasise this fact we write z_c in place of x_c.

The *abc* groups $\Gamma(a,b,c)$ are defined as follows: first we amalgamate G_a with $G_b \times \mathbb{Z}$ by identifying the centre of G_a with that of G_b, then we form the amalgamated free product of the resulting group with G_c by identifying the centre of the latter with the right-hand factor of $G_b \times \mathbb{Z}$. In symbols:

$$\Gamma(a,b,c) = G_a *_{z_a=z_b} (G_b \times \langle\, \zeta\,\rangle) *_{\zeta=z_c} G_c.$$

Theorem 3.1.8 *For all integers $1 \leqslant b \leqslant a < c$, the Dehn function of $\Gamma(a,b,c)$ is $\simeq n^{c+\frac{a}{b}}$.*

Variations on this construction yield other families of rational exponents [19].

By far the most comprehensive result concerning the structure of Dehn functions is due to Sapir, Birget and Rips. Their result, which we shall describe in a moment, essentially classifies the Dehn functions $\succeq n^4$. In particular they show that IP is dense in $[4,\infty)$.

Subsequently, Brady and Bridson [15] showed that Gromov's gap $(1,2)$ is the only gap in the isoperimetric spectrum:

Theorem 3.1.9 *For each pair of positive integers $p \geqslant q$, there exist finitely presented groups whose Dehn functions are $\simeq n^{2\alpha}$ where $\alpha = \log_2(2p/q)$.*

Corollary 3.1.10 *The closure of* IP *is* $\{1\} \cup [2,\infty)$.

Note that the exponents described in the above theorem are transcendental if they are not integers [80], Theorem 10.2. The easiest examples of groups as described in the above theorem are

$$G_{p,q} \;=\; \langle a,b,s,t \mid [a,b]=1, sa^q s^{-1} = a^p b,\; ta^q t^{-1} = a^p b^{-1}\rangle,$$

which we shall look at more closely in (7.2.12).

The Sapir-Birget-Rips Theorem. In [95] Mark Sapir, Jean-Camille Birget and Eliyahu Rips show that if a number $\alpha > 4$ is such that there is a constant $C > 0$ and a Turing machine that calculates the first m digits of the decimal expansion of α in time $\leqslant C2^{2^{Cm}}$, then $\alpha \in$ IP. Conversely, they show that if $\alpha \in$ IP then there is a Turing machine that calculates the first m digits of α in time $\leqslant C2^{2^{2^{Cm}}}$. (The discrepancy in the height of the two towers of exponentials is connected to the P = NP problem.) More generally they prove:

Theorem 3.1.11 *Let $\mathcal{D}_4$ be the set of $\simeq$ equivalence classes of Dehn functions $\delta(n) \succeq n^4$. Let $\mathcal{T}_4$ be the set of $\simeq$ classes of time functions $t(n) \succeq n^4$ of arbitrary Turing machines. Let $\mathcal{T}^4$ be the set of $\simeq$ classes of super-additive*[12] *functions that are fourth powers of time functions. Then $\mathcal{T}^4 \subseteq \mathcal{D}_4 \subseteq \mathcal{T}_4$.*

It is unknown whether $\mathcal{T}^4$ coincides with the $\simeq$ classes of all super-additive functions in $\mathcal{T}_4$. If it does, then the above theorem would completely classify Dehn functions $\succeq n^4$. In the light of Theorem 3.1.9, one suspects that Dehn functions $\succeq n^2$ are similarly unrestricted in nature.

As it stands, the above result already implies that any rational or other reasonable number, for example $\pi + e^2$, is the exponent of a Dehn function. Likewise, the following are Dehn functions: $2^{\sqrt{n}}$, e^{n^π}, $n^2\log_3(\log_7 n), \dots$

[12] $f(m+n) \geqslant f(n) + f(m)$ for all $n, m \in \mathbb{N}$

As one might guess from the statement, the theorem is proved by showing that one can encode the workings of a certain class of machines ("S-machines") into group presentations.

3.2 Examples of Large Dehn Functions

Thus far in this section I have concentrated on IP in order to explain the development of our understanding of Dehn functions. Let me offset this now by pointing out that many naturally occurring groups do not have Dehn functions that are bounded above by a polynomial function. We saw some such examples in (3.1.4). Here are some more simple examples of this type.

Consider the recursively-defined sequence of functions $\varepsilon_i(n) := 2^{\varepsilon_{i-1}(n)}$, where $\varepsilon_0(n) = n$ and $\varepsilon_1(n) = 2^n$. Let

$$B_m = \langle x_0, x_1, \ldots, x_m \mid x_i^{-1} x_{i-1} x_i = x_{i-1}^2 \text{ for } i = 1, \ldots, m\rangle. \qquad (3.2.1)$$

The best known of these groups is B_1, which has many manifestations, e.g. as a group of affine transformations of the real line, where x_0 acts as $t \mapsto t+1$ and x_1 as $t \mapsto 2t$.

Proposition 3.2.2 *The Dehn function of B_m is $\simeq \varepsilon_m(n)$.*

For the lower bound, see Exercise 7.2.11. The following exercises explain one method of establishing the upper bound.

Exercises 3.2.3 (i) Let w be a word in the generators of B_1. Show that one can transform w into a word of the form $x_1^m x_0^r x_1^{-m'}$ with $m, m' \geqslant 0$ by applying the defining relator $x_1^{-1} x_0 x_1 x_0^{-2}$ at most 2^n times. (Hint: Move each occurrence of x_1 in w to the left by replacing subwords $x_0 x_1$ with $x_1 x_0^2$, and $x_0^{-1} x_1$ with $x_1 x_0^{-2}$. Move all occurrences of x_1^{-1} to the right.)

(ii) Prove that $x_0 \in B_1$ has infinite order. (You could consider the representation $B_1 \to \mathrm{Aff}(\mathbb{R})$ described above.[13]) Deduce that $\delta_{B_m}(n) \preceq 2^n$. (Hint: By looking at the map $B_1 \to \langle x_1 \rangle$ that kills x_0 and the map $B_1 \to \mathbb{Z}_q \rtimes \mathbb{Z}$ that kills x_0^q (where q is an arbitrary odd prime), one can see that if $w = 1$ in Γ then the word obtained in (i) has $m = m'$ and $r = 0$.)

A less *ad hoc* proof of (ii) can be based on Britton's Lemma (see 7.2.4(ii) or [22], page 498):

(iii) Deduce from Britton's Lemma that if a word in the generators of B_m represents the identity and contains at least one occurrence of $x_m^{\pm 1}$ then it contains a subword of the form $w_0 = x_m^e w_1 x_m^{-e}$, where $e = \pm 1$ and w_1 is a word in the letters $\{x_i \mid i < m\}$ with $w_1 = x_{m-1}^p$ in B_{m-1}, where p is even if $e = 1$.

Arguing by induction on m, and a secondary induction on the number of occurrences of $x_m^{\pm 1}$ in w_0, show that one can replace w_0 by $x_{m-1}^{p/2}$ or x_{m-1}^{2p} by applying at most $\varepsilon_{m-1}(2p)$ relators from the presentation of B_{m-1}. Deduce that $\delta_{B_m}(n) \leqslant \varepsilon_m(n)$.

Example 3.2.4 Steve Gersten [45] showed that the Dehn function of the group

$$S = \langle x, y \mid (yxy^{-1})^{-1}x(yxy^{-1}) = x^2 \rangle$$

grows faster than any iterated exponential. Specifically, $\delta_S(n) \simeq \varepsilon_n(n)$. A classical theorem of Magnus states that all 1-relator groups have a solvable word problem. It is conjectured that $\varepsilon_n(n)$ is an upper bound on the Dehn functions of all 1-relator groups; in [46] Gersten established a weaker upper bound.

Exercise 3.2.5 Show that for every $m > 0$ there exists a monomorphism $B_m \to S$. (Hint: Conjugate (y^ixy^{-i}) by $(y^{i+1}xy^{-(i+1)})$.)

3.3 Groups of Classical Interest

In this subsection I shall describe what is known about the Dehn functions of various groups that are of interest for geometric reasons.

Low-Dimensional Topology. If S is a compact 2-manifold, then $\pi_1 S$ has a linear Dehn function unless S is a Torus or a Klein bottle, in which case $\pi_1 S$ has a quadratic Dehn function (see 3.1.1, 3.1.2, 1.3.5). The following theorem describes the situation for 3-dimensional manifolds – it follows easily from results of Epstein and Thurston [42] (cf. [17] and 7.1.4 below). Since all finitely presented groups arise as fundamental groups of closed n-manifolds for each $n \geqslant 4$ (see A.3.1), there can be no such general statement in higher dimensions.

Theorem 3.3.1 *Let M be a compact 3-manifold. Suppose that M satisfies Thurston's geometrization conjecture*[14]*.*

The Dehn function of $\pi_1 M$ is linear, quadratic, cubic, or exponential. It is linear if and only if $\pi_1 M$ does not contain $\mathbb{Z}^2$. It is quadratic if and only if $\pi_1 M$ contains $\mathbb{Z}^2$ but does not contain a subgroup $\mathbb{Z}^2 \rtimes_\phi \mathbb{Z}$ with $\phi \in \mathrm{GL}(2,\mathbb{Z})$ of infinite order. Subgroups $\mathbb{Z}^2 \rtimes_\phi \mathbb{Z}$ arise only if a finite-sheeted covering of M has a connected summand that is a torus bundle over the circle, and the Dehn function of $\pi_1 M$ is cubic only if each such summand is a quotient of the Heisenberg group (in which case ϕ is unipotent)[15]*.*

---

[13] More ambitiously, you could try to prove the following result of Higman, Neumann and Neumann (see [97] for a geometric treatment). Given a group $\Gamma = \langle \mathcal{A} \mid \mathcal{R} \rangle$ and an isomorphism $\phi : S_1 \to S_2$ between subgroups of Γ, one can form the HNN extension $\Gamma *_\phi = \langle \mathcal{A}, t \mid \mathcal{R},\ \phi'(s) = t^{-1}st,\ \forall s' \in S \rangle$, where $t \notin \mathcal{A}$, $S \subset F(\mathcal{A})$ is a set of words that maps bijectively to S_1, and for each $s \in S$ the word $\phi'(s) \in F(\mathcal{A})$ maps to $\phi(s) \in S_2$. Show that the map $\Gamma \to \Gamma *_\phi$ induced by $\mathrm{id}_{\mathcal{A}}$ is an injection.

[14] In the absence of this assumption it remains unknown whether every compact 3-manifold has a solvable word problem.

[15] $\pi_1 M$ has an exponential Dehn function if and only if M has a connected summand that is modelled on the geometry Sol – cf. 3.1.4

Remark 3.3.2 [Free-by-Cyclic Groups] If a 3-manifold M fibres over the circle then one sees from the long exact sequence in homotopy that $\pi_2 M = 0$ and that $\pi_1 M$ is a semi-direct product $\Sigma \rtimes_\phi \mathbb{Z}$, where Σ is the fundamental group of the surface fibre. Since $\pi_2 M = 0$, one knows that M does not split as a non-trivial connected sum, so the above theorem implies that if $\mathbb{Z}^2 \not\subseteq \Sigma$, then the Dehn function of $\pi_1 M$ is either linear or quadratic.

If M has boundary then Σ will be a finitely generated free group. Not all free-group automorphisms arise from fibrations of 3-manifolds, and it is has yet to be proved that the Dehn functions of arbitrary semi-direct products of the form $\Gamma = \Sigma \rtimes_\phi \mathbb{Z}$, with Σ free, are at most quadratic, cf. [69]. In [12] Bestvina and Feighn show that the Dehn function of Γ is linear if and only if $\mathbb{Z}^2 \not\subseteq \Gamma$.

There are strong analogies between mapping class groups of surfaces, Braid groups (more generally, Artin groups), and automorphism groups of free groups. These groups play important roles in low-dimensional topology. Bill Thurston proved that the Braid groups are automatic, [42] Chapter 9 (see also Charney [31]), and Lee Mosher proved that the mapping class groups of all surfaces of finite type are automatic [76]. As a consequence (see 6.3.2) we obtain:

Theorem 3.3.3 *The mapping class group of any surface of finite type satisfies a quadratic isoperimetric inequality.*

Hatcher and Vogtmann [58] and Gersten (unpublished) proved that the Dehn function of the group of (outer) automorphisms of any finitely generated free group is $\preceq 2^n$. Bridson and Vogtmann [24] proved that this bound is sharp in rank 3, and special considerations apply in rank 2.

Theorem 3.3.4 *Let F_r denote a free group of rank r. The Dehn function of* $\mathrm{Out}(F_2)$ *is linear. The Dehn function of* $\mathrm{Aut}(F_2)$ *is quadratic. The Dehn functions of* $\mathrm{Aut}(F_3)$ *and* $\mathrm{Out}(F_3)$ *are exponential. In general the Dehn functions of* $\mathrm{Aut}(F_r)$ *and* $\mathrm{Out}(F_r)$ *are* $\preceq 2^n$.

Lattices in Semisimple Lie Groups. Let G be a connected semisimple Lie group with finite centre and no compact factors. Associated to G one has a Riemannian symmetric space $X = G/K$, where $K \subseteq G$ is a maximal compact subgroup. A discrete subgroup $\Gamma \subset G$ is called a *lattice* if the quotient $\Gamma\backslash X$ has finite volume; the lattice is called *uniform* (or cocompact) if $\Gamma\backslash X$ is compact. The *rank* of G is the dimension of the maximal isometrically embedded flats $\mathbb{E}^r \hookrightarrow X$.

If G has rank 1 then X has strictly negative curvature (e.g. $G = \mathrm{SO}(n,1)$ and $X = \mathbb{H}^n$) and in general (e.g. $G = \mathrm{SL}(n,\mathbb{R})$) X has non-positive curvature (see, for example, [22] Chapter II.10). It follows that the Dehn functions of uniform lattices are linear (in the rank 1 case) or quadratic (the higher rank case) – see Section 6.

Each non-uniform lattice in a rank 1 group contains non-trivial subgroups that stabilize points at infinity in the symmetric space X; these subgroups leave invariant the horospheres centred at the fixed points at infinity. We use the term *horospherical* to describe these subgroups. An example of a horospherical subgroup is the fundamental group of the boundary torus in a hyperbolic knot complement. Each maximal horospherical subgroup contains a nilpotent subgroup of finite index: in the case $G = \mathrm{SO}(n,1)$, this nilpotent subgroup is isomorphic to $\mathbb{Z}^{n-1}$, and in the case $G = \mathrm{SU}(n,1)$ it is isomorphic to H_{2n-1}, the integer Heisenberg group.

Theorem 3.3.5 *Let G be a semisimple Lie group of rank 1 and let $\Gamma \subset G$ be a lattice. If Γ is uniform then its Dehn function is linear. If Γ is non-uniform then its Dehn function is equal to that of each of its maximal horospherical subgroups.*

This result is due to Gromov [56].

Example 3.3.6 It follows from our discussion in 2.2 that non-uniform lattices in $\mathrm{SU}(2,1)$ have cubic Dehn functions, whereas those in $\mathrm{SU}(n,1)$ with $n > 2$ have quadratic Dehn functions. More generally, it follows from the above theorem that a non-uniform lattice in a rank 1 group G will have a quadratic Dehn function unless the symmetric space for G is the hyperbolic plane over the real, complex, quaternionic or Cayley numbers. For the real hyperbolic plane the Dehn function of non-uniform lattices is linear (3.1.1), in the complex case ($G = \mathrm{SU}(2,1)$) it is cubic, and it is also believed to be cubic in the remaining cases.

The following theorem of Leuzinger and Pittet [62], which builds on the work of Gromov on solvable groups [56], completes the picture of Dehn functions for lattices in rank 2.

Theorem 3.3.7 *If G is a connected semisimple Lie group with finite centre and rank 2, then the Dehn function of any irreducible, non-uniform lattice in G is $\simeq 2^n$.*

The situation for non-uniform lattices in rank $\geqslant 3$ is more complicated and is the subject of active research. We refer the reader to Gromov [56] for an exciting glimpse of some of the issues that arise and to Druţu [38] and Leuzinger-Pittet [63] for significant recent progress in this direction. The following assertion of Bill Thurston illustrates some of the subtleties involved in higher rank: *Dehn function of $\mathrm{SL}(3,\mathbb{Z})$ is exponential, but the Dehn function of $\mathrm{SL}(n,\mathbb{Z})$ is quadratic if $n > 3$.*

See [42] page 230 for a proof of this statement in the case $n = 3$ (cf. [38] and [56] page 91). A complete proof is not available in the case $n > 3$. Druţu's recent work has helped to clarify the situation, but there remains much work to be done in this direction.

Nilpotent Groups. We saw in (3.1.2) that abelian groups satisfy a quadratic isoperimetric inequality. Using a modest amount of knowledge about the structure of nilpotent groups, it is not hard to show that all finitely generated nilpotent groups satisfy a polynomial isoperimetric inequality (see [56] for example). But determining the degree of the optimal bound on the Dehn function, both in general and for specific examples, is a more delicate matter, as our earlier discussion of the Heisenberg groups illustrates.

Gromov, [56] Chapter 5, gives an enticing overview of this area. In particular he sketches a reason why nilpotent groups of class c should have Dehn functions that are polynomial of degree $\leqslant c+1$ and gives a proof of this inequality for groups where the Lie algebra of the Malcev completion is graded. (For a detailed account of this last result, and extensions, see Pittet [89].) A number of other researchers have obtained related results using both geometric and combinatorial methods. In particular, Hidber [59] gives a purely algebraic proof that the Dehn function of a nilpotent group of class c is bounded above by a polynomial of degree $2c$.

Finally, I should mention that the study of Dehn functions of non-nilpotent solvable groups is also an active area of research. Indeed this is closely connected to the study of Dehn functions for higher-rank lattices.

Let me end this brief survey of our knowledge of Dehn functions for specific groups by making it clear that I have omitted far more than I have included. I apologise to the many colleagues whose excellent work I have been forced to ignore by reason of space and time.

3.4 Dehn Functions of Products

The following exercises describe how Dehn functions behave under the formation of products. Their behaviour under more complicated operations such as amalgamated free products, HNN extensions, and central extensions is less straightforward.

Exercises 3.4.1 (i) A subgroup H of a group G is called a *retract* if there is a homomorphism $G \to H$ whose restriction to H is the identity. Show that if H is a retract of the finitely presented group G, then H is finitely presented and $\delta_H(n) \preceq \delta_G(n)$. (Hint: First note that H is finitely generated. Take a finite subset that generates H and argue that it can be extended to a finite generating set for G by adding elements k of the kernel of $G \to H$. Argue that one can take a finite presentation for G with this generating set. Add the relations $k = 1$.)

(ii) Let G_1 and G_2 be infinite, finitely presented groups. Show that the Dehn function of $G_1 \times G_2$ is $\simeq \max\{n^2, \delta_{G_1}(n), \delta_{G_2}(n)\}$, and that that of the free product $G_1 * G_2$ is $\simeq \max\{\delta_{G_1}(n), \delta_{G_2}(n)\}$. (Use (i) for the bounds $\succeq$.)

4 Van Kampen Diagrams

Let $\langle \mathcal{A} \mid \mathcal{R} \rangle$ be a finite presentation of a group Γ and let w be a word in the letters $\mathcal{A}^{\pm 1}$. Suppose that $w = 1$ in Γ. Roughly speaking, a van Kampen diagram for w is a planar CW complex that portrays a scheme for reducing w to the empty word by applying a sequence of relations $r \in \mathcal{R}$; the number of 2-cells in the diagram is the number of relations that one applies and is therefore at least as great as $\text{Area}_a(w)$, as defined in (1.2.2). Conversely, we shall see that one can always construct a van Kampen diagram for w that has $\text{Area}_a(w)$ 2-cells. It follows that the Dehn function of $\langle \mathcal{A} \mid \mathcal{R} \rangle$ can be interpreted in terms of isoperimetric inequalities for planar diagrams.

Max Dehn was the first to use planar diagrams in order to study word problems [34], but his diagrams arose in concrete settings (primarily as regions in a tessellated hyperbolic plane). The idea of using diagrams to study relations in arbitrary finitely presented groups is due to E. van Kampen [61]. The idea was rediscovered by Roger Lyndon in the 1960s. At about the same time C. Weinbaum brought van Kampen's original paper to light and made interesting applications of it.

There are a number of correct proofs of the celebrated van Kampen Lemma in the literature. The use of pictures in these proofs causes disquiet in some circles, so I have tried to fashion the following proof in a manner that will allay such misgivings.

4.1 Singular Disc Diagrams

Fix an orientation on $\mathbb{R}^2$. A *singular disc diagram* D is a compact, contractible subset of the plane endowed with the structure of a finite combinatorial 2-complex. (See Appendix A for basic definitions concerning combinatorial complexes.)

We write $\text{Area}_c\, D$ to denote the number of 2-cells in D. And given a vertex $p \in D$ we write $\text{Diam}_p\, D$ to denote the maximum of the distance from p to the other vertices $v \in D$, where "distance" is the number of 1-cells traversed by a shortest path joining p to v in the 1-skeleton of D.

To avoid pathologies, we assume the 1-cells $e : [0,1] \to D \hookrightarrow \mathbb{R}^2$ are smoothly embedded. Associated to each 1-cell one has two *directed edges* $\varepsilon(t) = e(t)$ and $\varepsilon(t) = e(1-t)$. Let $\mathcal{A}_D$ denote the set of directed edges. (By definition $\varepsilon = \varepsilon$.)

The *boundary cycle* of D is the loop of directed edges describing the frontier of the metric completion of $\mathbb{R}^2 \smallsetminus D$ in the positive (anti-clockwise) direction – it consists of a *thin part*, where the underlying 1-cells do not lie in the boundary of any 2-cell, and a *thick part*; the boundary cycle traverses each 1-cell in the thick part once and each 1-cell in the thin part twice.

Definition 4.1.1 [Labelled Diagrams] *Let $\mathcal{A}$ be a set and let $\mathcal{A}^{-1}$ be the set of symbols $\{a^{-1} \mid a \in \mathcal{A}\}$. A diagram over $\mathcal{A}$ consists of a singular disc diagram D and a (labelling) map $\lambda : \mathcal{A}_D \to \mathcal{A} \cup \mathcal{A}^{-1}$ such that $\lambda(\varepsilon) = \lambda(\varepsilon)^{-1}$ for all $\varepsilon \in \mathcal{A}_D$.*

λ extends to a map from the set of directed edge-paths in D to the set of words in the letters $\mathcal{A} \cup \mathcal{A}^{-1}$. The face labels of D are the words that this map assigns to the attaching loops of the 2-cells of D (beginning at any vertex and proceeding with either orientation).

Proposition 4.1.2 *Let $\mathcal{A}$ be a set, let D be a diagram over $\mathcal{A}$ and let $\mathcal{R}_*$ be a set of words that contains the face labels of D. If a word w occurs as the label* *on the boundary cycle of D, read from some vertex p in the boundary of D, then in the free group $F(\mathcal{A})$*

$$w = \prod_{i=1}^{\alpha} x_i^{-1} r_i x_i,$$

where $\alpha = \mathrm{Area}_c\, D$, the words x_i have length $|x_i| \leqslant \mathrm{Diam}_p\, D$, and $r_i \in \mathcal{R}_*$. In particular $w = 1$ in the group $\langle \mathcal{A} \mid \mathcal{R}_* \rangle$.

Proof Fix D and p. In the 1-skeleton of D we choose a geodesic spanning tree T rooted at p (see Exercise 4.1.3).

Arguing by induction (the base step is trivial) we may assume that the proposition has been proved for diagrams D' with $\mathrm{Area}_c\, D' < \mathrm{Area}_c\, D$ and for diagrams with $\mathrm{Area}_c\, D' = \mathrm{Area}_c\, D$ where D' has fewer 1-cells than D.

We say that D has a dangling edge if it has a vertex other than p that has only one edge incident at it. If D has such an edge then we may apply our inductive hypothesis to the diagram obtained by removing it – the resulting diagram has the same area as D, its diameter is no greater than that of D, and its boundary label is obtained from that of D by free reduction. Thus we may assume that D has no dangling edges.

If D were a tree it would have dangling edges (or be a single point). Thus $D \neq T$. We follow the boundary cycle of D from p until we encounter the first directed edge ε that is not in T; let a be the label on ε, let w_1 be the label on the segment of the boundary cycle that precedes ε and let w_2 be the label on the segment that follows it. The part of the boundary cycle labelled w_1 is an injective path, because it lies entirely in the tree T and must be locally injective since a backtracking would imply that D had a dangling edge. In particular w_1 has length at most $\mathrm{Diam}_p\, D$.

Since T contains all of the vertices of D, we do not disconnect D by removing the open 1-cell underlying ε, and hence this 1-cell must lie in the boundary of some 2-cell E. Suppose that the attaching loop of E (read in the positive direction from the initial vertex of ε) has label $r^{-1} := au^{-1}$.

Consider the subcomplex D' obtained from D by deleting the open 1-cell labelled ε and the interior of E. Note that D' is again a diagram over $\mathcal{A}$ (its labelling map is just the restriction of the labelling map of D), its set of face labels is a subset of the face labels of D, its diameter is the same as that of D (because the geodesic spanning tree T is entirely contained in D') and its boundary cycle, read from p, is $w' := w_1 u w_2$. In the free group $F(\mathcal{A})$ we have

$$w' = (w_1 r w_1^{-1})(w_1 a w_2) = (w_1 r w_1^{-1}) w.$$

We have argued that $|w_1| \leqslant \mathrm{Diam}_p\, D = \mathrm{Diam}_p\, D'$. And by induction we may assume that w' can be expressed as a product of conjugates of at most $\mathrm{Area}_c\, D' = \mathrm{Area}_c\, D - 1$ face labels, with conjugating elements of length at most $\mathrm{Diam}_p\, D' = \mathrm{Diam}_p\, D$. This completes the induction. $\square$

One can give a shorter proof of the above proposition if one ignores the length of the conjugating elements x_i; this weaker form of the result is more standard, e.g. [66].

> *Exercise 4.1.3* Let $\mathcal{G}$ be a connected graph (1-dimensional CW complex). Let d be a length metric in which each edge has length 1. Fix a vertex $p \in \mathcal{G}$. Prove that $\mathcal{G}$ contains a geodesic spanning tree rooted at p, i.e. a 1-connected subgraph T that contains a path of length $d(p, v)$ from p to each vertex $v \in \mathcal{G}$.

4.2 Van Kampen's Lemma

Definition 4.2.1 [Van Kampen Diagrams] *Let $\mathcal{A}$ be a set, let $\mathcal{R}$ be a set of words in the letters $\mathcal{A}^{\pm 1}$ and let $\mathcal{R}_*$ be the smallest set of words that contains $\mathcal{R}$ and is closed under the operations of taking cyclic permutations and inverses of words. (Note that $\langle \mathcal{A} \mid \mathcal{R} \rangle \cong \langle \mathcal{A} \mid \mathcal{R}_* \rangle$.)*

If w, D and p are as in the above proposition, then D is called a van Kampen diagram *for w over $\langle \mathcal{A} \mid \mathcal{R} \rangle$ with basepoint p.*

Theorem 4.2.2 (Van Kampen's Lemma) *Let $\mathcal{A}$ be a set, let w be a word in the letters $\mathcal{A} \cup \mathcal{A}^{-1}$, and let $\mathcal{R}$ be a set of words in these letters.*

(1) *$w = 1$ in the group $\Gamma = \langle \mathcal{A} \mid \mathcal{R} \rangle$ if and only if there exists a van Kampen diagram for w over $\langle \mathcal{A} \mid \mathcal{R} \rangle$.*
(2) *If $w = 1$ in Γ then*

$$\mathrm{Area}_a(w) = \min\{\mathrm{Area}_c\, D \mid D \text{ a van Kampen diagram for } w \text{ over } \langle \mathcal{A} \mid \mathcal{R} \rangle\}.$$

In order to complete the proof of this theorem we shall need two lemmas. In the first we consider the following ordering on diagrams over $\mathcal{A}$ that have an initial vertex[16] specified in the boundary cycle: $D \prec D'$ if D' has fewer 1-cells than D and the words labelling the boundary cycles of D and D', read from their initial vertices, are equal as elements of the free group $F(\mathcal{A})$.

Lemma 4.2.3 *If D, with initial point p, is minimal in the ordering $\prec$, then the boundary label of D is a freely reduced word.*

Proof We shall assume that D is a diagram whose boundary label w is not freely reduced and construct a diagram $\prec D$.

Since w is not reduced, there is a pair of successive directed edges $\varepsilon, \varepsilon'$ in the boundary cycle that are labelled a, a^{-1} respectively, where $a \in \mathcal{A} \cup \mathcal{A}^{-1}$. If the initial vertex of ε is equal to the terminal vertex of ε' then we can delete

[16] A choice of "initial vertex" includes the specification of which edge of the boundary cycle is to be traversed first. Nevertheless, when no confusion is threatened, one talks as if the "initial vertex" is simply a vertex of D.

from D these edges together with the contractible region that they enclose, thus obtaining a diagram $\prec D$.

If the initial vertex of ε is not equal to the terminal vertex of ε' then[17] we can connect the latter vertex to the former by a smooth arc $c : [-1,1] \to \mathbb{R}^2$ that intersects D only at its endpoints. Let $T \subset \mathbb{R}^2$ be the open disc enclosed by the loop $\varepsilon\varepsilon' c$; we shall collapse T in a controlled manner. Let $\Delta = \{(x,y) \mid 1 > y > |x|,\, |x| < 1\} \subset \mathbb{R}^2$ and fix a diffeomorphism $\phi : \Delta \to T$ that has a continuous extension to Δ with $\phi|_{[-1,1]\times\{1\}} = c$ and $\phi(-t,t) = \varepsilon'(1-t)$ and $\phi(t,t) = \varepsilon(t)$ for all $t \in [0,1]$. The map $T \to \{0\} \times \mathbb{R}$ that sends $z = \phi(x,y)$ to y has a continuous extension $\pi : \mathbb{R}^2 \to \mathbb{R}^2$ that is a diffeomorphism on the complement of the closure of T.

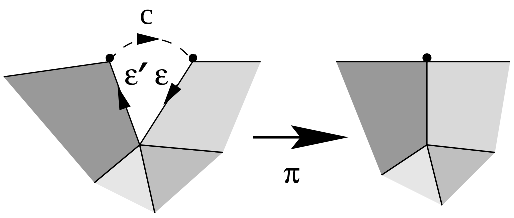

Figure 4.2.4 *Reducing the boundary label*

$\overline{D} = \pi(D)$ inherits a combinatorial structure from D as well as a choice of initial point for its boundary cycle. $\overline{D}$ has fewer 1-cells than D because $\pi \circ \varepsilon^{-1} = \pi \circ \varepsilon'$. The directed edges $\pi \circ \varepsilon_i$ of $\overline{D}$ inherit the labelling $\lambda(\varepsilon_i)$ from D, and the label on the boundary cycle of $\overline{D}$, read from its initial point, is obtained from w by deleting the subword aa^{-1} corresponding to $\varepsilon\varepsilon'$. Thus $\overline{D} \prec D$. □

Remark 4.2.5 If one employs a suitably natural procedure for choosing the edge ε, then the proof given above actually constitutes an algorithm for transforming a diagram D whose boundary label is not freely reduced into a diagram $D' \prec D$. By repeated application of this algorithm one obtains a diagram $D_0 \prec D$ whose boundary label is reduced. Moreover, the set of face labels of D_0 is contained in the set of face labels of D, and $\mathrm{Area}_c\, D_0 \leqslant \mathrm{Area}_c\, D$.

The following lemma is used to pass from diagrams whose boundary labels are reduced to those whose labels are not.

Lemma 4.2.6 *Let $\mathcal{A}$ be a set, let w be a word in the letters $\mathcal{A} \cup \mathcal{A}^{\pm 1}$ and let w_0 be the reduced word that is equal to w in $F(\mathcal{A})$. Given a diagram D_0 for w_0 over $\mathcal{A}$, one can construct a diagram D for w with* $\mathrm{Area}_c\, D_0 = \mathrm{Area}_c\, D$ *so that the set of face labels of D is the same as that of D_0.*

[17] There are no hidden assumptions here: ε and ε' may be in the thin part of the boundary or in the thick part, and one of them might be a loop.

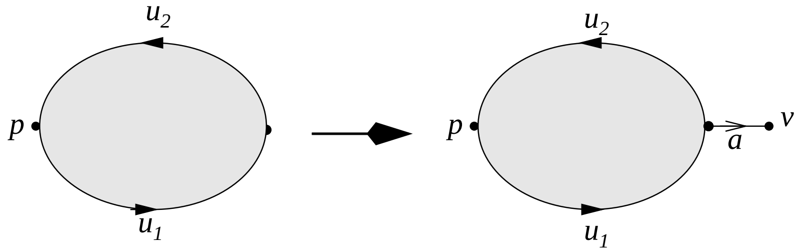

Figure 4.2.7 *Changing the boundary label from u_1u_2 to $u_1aa^{-1}u_2$.*

Proof w is obtained from w_0 by repeatedly inserting pairs of letters aa^{-1} with $a \in \mathcal{A}$. To modify the boundary label of a diagram by such an insertion, one adds a new vertex v of valence 1 and an edge labelled a to v from the appropriate vertex of the boundary cycle (figure 4.2.7).

The Proof of Van Kampen's Lemma. If $w = 1$ in $\Gamma = \langle \mathcal{A} \mid \mathcal{R} \rangle$ then in the free group $F(\mathcal{A})$ we have:

$$w \overset{\text{free}}{=} \prod_{i=1}^{N} x_i^{-1} r_i x_i$$

where $r_i \in \mathcal{R}^{\pm 1}$ and $N = \mathrm{Area}_a w$. The word W on the right of this equality is the boundary label on the "lollipop" diagram D_1 shown in figure 4.2.8; note that $\mathrm{Area}_c\, D_1 = N$.

Let $D_0 \preceq D_1$ be a $\prec$-minimal diagram. The boundary label of D_0 is the freely reduced word w_0 that is equal to w in $F(\mathcal{A})$, the face labels of D_0 are a subset of those of D_1, and $\mathrm{Area}_c\, D_0 \leqslant \mathrm{Area}_c\, D_1 = \mathrm{Area}_a w$ (Lemma 4.2.3 and (4.2.5)). By applying Lemma 4.2.6 to D_0 we obtain a van Kampen diagram D of area $\leqslant N$ for w over $\langle \mathcal{A} \mid \mathcal{R} \rangle$. This proves the implication "only if" in (1) and the inequality $\geqslant$ in (2). Proposition 4.1.2 provides the complementary "if" implication and $\leqslant$ inequality. □

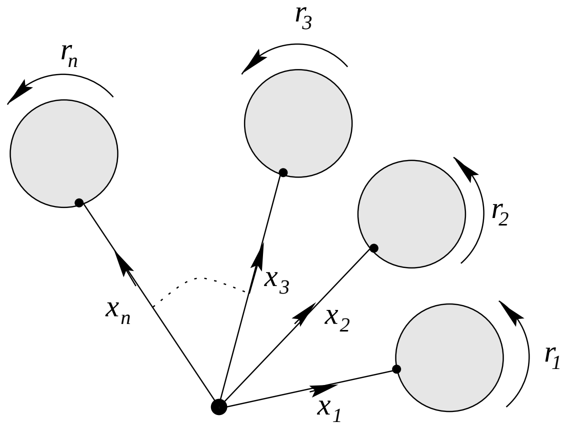

Figure 4.2.8 *The lollipop diagram*

4.3 Words and Van Kampen Diagrams as Maps

In this subsection I shall assume that the reader is familiar with the material in Section A.2 (Appendix A).

If D is a van Kampen diagram over $\langle \mathcal{A} \mid \mathcal{R}\rangle$ with basepoint p, then there is a unique label-preserving combinatorial map from the 1-skeleton of D to the Cayley graph $\mathcal{C}_{\mathcal{A}}(\Gamma)$ that sends p to the vertex $1 \in \Gamma$. This extends to a combinatorial map from D to the universal covering $\tilde{K}$ of the standard 2-complex $K(\mathcal{A}:\mathcal{R})$.

Let M be a closed, smooth Riemannian manifold and let $\Gamma = \langle \mathcal{A} \mid \mathcal{R}\rangle$ be a finite presentation of the fundamental group of M. Let $\tilde{K}$ be the universal covering of $K(\mathcal{A}:\mathcal{R})$. We fix a basepoint $p \in \tilde{M}$, and for every $a \in \mathcal{A}$ we choose a geodesic c_a joining p to $a.p$. These choices give rise to a Γ-equivariant map from $\mathcal{C}_{\mathcal{A}}(\Gamma) = \tilde{K}^{(1)}$ to $\tilde{M}$: this map sends the 1-cell labelled a emanating from γ homeomorphically onto the segment $\gamma.c_a$. Since $\tilde{M}$ is simply-connected, we may extend this map across the 2-cells of $\tilde{K}$ in a Γ-equivariant manner. We choose this extension so that on each 2-cell it is smooth almost everywhere and has finite area.

If w is a word in the letters $\mathcal{A} \cup \mathcal{A}^{-1}$, then for each $\gamma \in \Gamma$ there is a unique edge-path in $\mathcal{C}_{\mathcal{A}}(\Gamma) = \tilde{K}^{(1)}$ that begins at γ and is labelled w. We write $\hat{w}^{\gamma}$ to denote the image of this path in $\tilde{M}$ (except that if $\gamma = 1$ we write $\hat{w}$ instead of $\hat{w}^1$). Such paths in $\tilde{M}$ are called *word-like*.

If D is a van Kampen diagram for w over $\langle \mathcal{A} \mid \mathcal{R}\rangle$, then by composing the above maps $D \to \tilde{K}$ and $\tilde{K} \to \tilde{M}$ we obtain a map $h_D : D \to \tilde{M}$ whose restriction to the boundary cycle of D is a parameterization of the loop $\hat{w}$.

5 The Equivalence $\mathrm{Fill}_0^M \simeq \delta_{\pi_1 M}$

This section is devoted entirely to the proof of the Filling Theorem:

Theorem 5.0.1 *The 2-dimensional, genus-zero isoperimetric function* Fill_0^M *of any smooth, closed Riemannian manifold* M *is* $\simeq$ *equivalent to the Dehn function* $\delta_{\pi_1 M}$ *of the fundamental group of* M*.*

5.1 The Bound $\mathrm{Fill}_0^M \preceq \delta_{\pi_1 M}$

This direction of the proof is substantially easier than the other. In order to understand the proof, the reader will need to have absorbed the definition of a van Kampen diagram.

Proposition 5.1.1 *If* M *is a smooth, closed Riemannian manifold then* $\Gamma := \pi_1 M$ *is finitely presented and* $\mathrm{Fill}_0^M \preceq \delta_\Gamma$*.*

Proof Corollary A.4.2 of the Appendix shows that Γ is finitely presented. We fix a finite presentation for Γ and assume that the universal cover $\tilde{K}$ of the standard 2-complex of this presentation has been mapped to $\tilde{M}$ as explained in the preceding subsection. We identify Γ (the 0-skeleton of $\tilde{K}$) with its image in $\tilde{M}$. We define λ to be the maximum distance of any point of $\tilde{M}$ from Γ, we define μ to be the maximum of the lengths of the 1-cells of $\tilde{K}$, as measured in $\tilde{M}$, and we define $m = \max\{d_\Gamma(\gamma,\gamma') \mid d_{\tilde{M}}(\gamma,\gamma') \leqslant 2\lambda+1\}$, where d_Γ is the word metric associated to our chosen generators for Γ.

The images in $\tilde{M}$ of the 2-cells of $\tilde{K}$ are discs of finite area; let α be the maximum of these areas.

Let w be a word in the given generators that equals $1 \in \Gamma$ and consider the corresponding piecewise-geodesic loop $\hat{w}$ in $\tilde{M}$. Choose a van Kampen diagram D for w with $\mathrm{Area}_a(w)$ 2-cells, and consider the associated map $h_D : D \to \tilde{M}$, which fills $\hat{w}$. The area of this map is at most α times the number of 2-cells in D, hence

$$\mathrm{FArea}(\hat{w}) \leqslant \alpha \mathrm{Area}_a(w) \leqslant \alpha\, \delta_\Gamma(|w|).$$

Given a loop $c : \mathbb{S}^1 \to \tilde{M}$ of finite length $l(c)$, parameterized by arc length, we choose a set of n equally-spaced points $\theta_0, \dots, \theta_{n-1} \in \mathbb{S}^1$, where n is the least integer greater than $l(c)$. We then choose a geodesic segment σ_i from each $c(\theta_i)$ to a nearest point $\gamma_i \in \Gamma \subset \tilde{M}$. The distance in $\tilde{M}$ between successive γ_i (indices mod n) is at most $2\lambda+1$ and hence γ_i can be connected to γ_{i+1} by a word-like path $\hat{u_i}^{\gamma_i}$ of length at most $m\mu$, where u_i is a word of (algebraic) length m. Since each of the loops[18] $\sigma_i \hat{u}_i^{\gamma_i} \overline{\sigma}_{i+1} \overline{c}|_{[\theta_i,\theta_{i+1}]}$ has length at most $L := m\mu+1+2\lambda$, we have

$$\mathrm{FArea}(c) \leqslant \mathrm{FArea}(\hat{U_c}) + n\,\mathrm{Fill}_0^M(L),$$

where U_c is the concatenation of the words u_i (see figure 5.1.2).

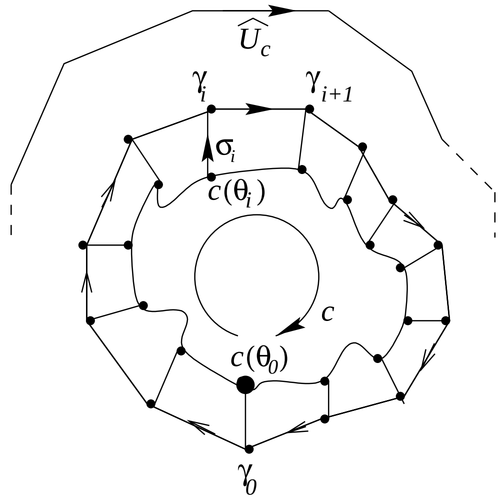

Figure 5.1.2 *Approximating c by the word-like loop $\widehat{U}_c$*

[18] an overbar denotes reversed orientation

The loop c is arbitrary, the word U_c has (algebraic) length at most $nm\mu$, and $n \leqslant l(c)+1$. Thus, the two inequalities displayed above imply that

$$\mathrm{Fill}_0^M(l) \ \leqslant\ \alpha\delta_\Gamma\big(m(l+1)\big) + (l+1)\,\mathrm{Fill}_0^M(L).$$

for all $l > 0$. In particular, since $\mathrm{Fill}_0^M(L)$ is a constant, $\mathrm{Fill}_0^M \preceq \delta_\Gamma$.

5.2 The Bound $\delta_{\pi_1 M} \preceq \mathrm{Fill}_0^M$.

There are many subtleties concerning the nature of solutions to Plateau's problem in Riemannian manifolds, but the existence of least-area discs (although highly non-trivial) has little to do with the fine structure of the spaces concerned. Indeed Igor Nikolaev [78] showed that one can solve Plateau's problem in any complete simply-connected geodesic space with an upper curvature[19] bound k. In this generality, when endowed with the pull-back path metric, a least-area spanning disc will itself have curvature $\leqslant k$. (To get some intuition about why this is true, observe that if a disc embedded in Euclidean 3-space has a point of positive curvature, then there is an obvious local pushing move that reduces the area of the disc without disturbing its boundary.)

Definition 5.2.1 *Let D be a metric space homeomorphic to a (perhaps singular[20]) 2-disc and fix $\varepsilon > 0$. A set $\Sigma \subset D$ is said to ε-fill D if every point of D is a distance less than ε from Σ and every point of the boundary cycle ∂D can be connected to a point of $\partial D \cap \Sigma$ by an arc in ∂D that has length at most ε.*

The only fact that we need concerning the nature of solutions to Plateau's problem is that loops in the universal covering of a closed Riemannian manifold can be filled by discs that exhibit the following crude consequence of the curvature bound described above.

Proposition 5.2.2 *If M is a complete Riemannian manifold of curvature $\leqslant k$, then the induced metric on every least-area disc $D \to \tilde{M}$ is such that D can be ρ_k-filled by a set of cardinality less than $\lambda_k(\mathrm{Area}(D) + |\partial D| + 1)$, where $|\partial D|$ denotes the length of the boundary of D and the constants λ_k and ρ_k depend only on k.*

Proof When equipped with the pull-back metric D has curvature $\leqslant k$. In the Riemannian setting this means that if a metric ball of radius $r < \pi/(2\sqrt{k})$ is contained in the interior of D, then the area of that ball is at least as great as the area of a disc of radius r in M_k^2. So if $a(k,r)$ denotes the area of such a disc, then there can be at most $\mathrm{Area}(D)/a(k,r)$ disjoint balls of radius r in the interior of D.

[19] In the sense of A.D. Alexandrov; see Appendix B for the definition. In Nikolaev's theorem there is a natural restriction on the length of the loops being filled if $k > 0$.

[20] A singular disc is a space homeomorphic to the underlying space of a singular disc diagram, as defined in (4.1). In the Riemannian setting (dimension $\geqslant 3$) one can avoid the need to discuss (topologically) singular discs by considering fillings of embedded loops only (cf. 2.1.5).

We set $r = r_k := \pi/(4\sqrt{k})$ and choose a maximal collection of disjoint balls of radius r in the interior of D. Let Σ_0 denote the set of centres of these balls. We also choose a collection Σ_1 of no more than $\frac{1}{r_k}|\partial D| + 1$ points along ∂D so that every point of ∂D can be connected to a point in Σ_1 by an arc of length less than r_k. By construction, the balls of radius $2r_k$ centred at the points $\Sigma_0 \cup \Sigma_1$ cover D and the cardinality of $\Sigma_0 \cup \Sigma_1$ is bounded above by $\frac{1}{a(k,r_k)}\mathrm{Area}(D) + \frac{1}{r_k}|\partial D| + 1$. Set $\rho_k = 2r_k$ and $\lambda_k = \max\{\frac{1}{a(k,r_k)}, \frac{1}{r_k}\}$. $\square$

In order to establish the reverse inequality in the Filling Theorem we shall use the following technical tool for manufacturing combinatorial discs out of ε-filling sets.

5.2.3 Cellulation Lemma. *Let D be a length space homeomorphic to a (perhaps singular) 2-disc, and suppose that D is ε-filled by a set Σ of cardinality N. Then there exists a combinatorial 2-complex Φ, homeomorphic to the standard 2-disc, and a continuous map $\phi : \Phi \to D$ such that:*

(1) *Φ has less than $8N$ faces (2-cells) and each is a k-gon with $k \leqslant 12$;*
(2) *the restriction of ϕ to each 1-cell in Φ is a path of length at most 2ε;*
(3) *$\phi|_{\partial\Phi}$ is a monotone parameterisation of ∂D and $\Sigma \cap \partial D$ lies in the image of the 0-skeleton of $\partial\Phi$.*

In the case of the ε-fillings yielded by Proposition 5.2.2 (which are the focus of our concern), instead of using the decomposition of D furnished by the Cellulation Lemma, one might use the dual to the Voronoi decomposition for the given filling — this dual will generically be a triangulation (cf. [99] 5.58). Some care is needed in pursuing this remark, but nevertheless we use it as a pretext[21] for relegating the proof of the Cellulation Lemma to Appendix C.

The Remainder of the Proof of the Filling Theorem.

It remains to show that $\delta_\Gamma \preceq \mathrm{Fill}_0^M$, where $\Gamma = \pi_1 M$. Let $k > 0$ be an upper bound on the sectional curvature of M.

We fix a basepoint $p \in \tilde{M}$ and choose a number $\rho > 0$ sufficiently large to ensure that the balls of radius $\rho/8$ about $\{\gamma.p \mid \gamma \in \Gamma\}$ cover $\tilde{M}$ and that $\rho > 8\rho_k$ (notation of 5.2.2). Let $\mathcal{A}$ be the set of $a \in \Gamma$ such that $d(a.p, p) < \rho$ and let $\mathcal{R}$ be the set of words in the symbols $\mathcal{A} \cup \mathcal{A}^{-1}$ that have length $\leqslant 12$ and equal the identity in Γ. (Note that $\mathcal{A}$ contains a letter that represents $1 \in \Gamma$.) Corollary A.4.2 shows that $\langle \mathcal{A} \mid \mathcal{R} \rangle$ is a presentation of Γ. We shall show that every null-homotopic word w over this presentation satisfies

$$\mathrm{Area}_a(w) \leqslant 4\lambda_k \left(\mathrm{Fill}_0^M(\rho\,|w|) + \rho\,|w| + 1\right).$$

Given a word w with $w = 1$ in Γ we consider the piecewise geodesic loop $\hat{w}$ in $\tilde{M}$ (notation of 4.3). This loop has length less than $\rho|w|$ and hence[22]

[21] The honest reason for this deferral is that the proof is lengthy and inelegant.

[22] If one wants to quote Morrey directly here one should perturb $\hat{w}$ to ensure that it is embedded.

can be filled with a least-area disc $f : D \to \tilde{M}$ of area at most $\mathrm{Fill}_0^M(\rho\,|w|)$. Using Proposition 5.2.2 we can ρ_k-fill D with a set Σ of cardinality less than $N := \lambda_k(\mathrm{Fill}_0^M(\rho\,|w|) + \rho\,|w| + 1)$. Increasing the cardinality of Σ by at most $|w|$, we may assume that it contains the vertices of $\hat{w}$.

Consider a combinatorial 2-disc Φ and a map $\phi : \Phi \to D$ as furnished by the Cellulation Lemma. Our aim is to label Φ so that it becomes a van Kampen diagram for w over $\langle \mathcal{A} \mid \mathcal{R} \rangle$. The composition $f \circ \phi : \Phi \to \tilde{M}$ will guide us in this construction. Note that the restriction of $f \circ \phi$ to $\partial\Phi$ is a monotone parameterization of $\hat{w}$. The initial point of $\hat{w}$ determines a basepoint for Φ.

For each vertex v in the interior of Φ we choose a point $\gamma_v \cdot p$ in the Γ orbit of p that is closest to $f \circ \phi(v)$. If v and v' are the vertices of a 1-cell in Φ, then $f \circ \phi(v)$ and $f \circ \phi(v')$ are a distance at most $2\rho_k$ apart in $\tilde{M}$ (the second property of Φ in the Cellulation Lemma). It follows that $d(\gamma_v \cdot p, \gamma_{v'} \cdot p) \leqslant 2\rho_k + \rho/4$, which is less than $\rho/2$. Hence there exists a generator $a \in \mathcal{A}$ such that $a = \gamma_v^{-1}\gamma_{v'}$ in Γ. We introduce the label a on the edge in Φ joining v to v'.

Among the vertices of $\partial\Phi$ we have a set of distinguished vertices, namely those mapping to the vertices of $\hat{w}$. Call these $x_0, \ldots, x_{n-1}$, corresponding to the vertices $w_i \cdot p$ on $\hat{w}$, where w_i is the i-th prefix of w.

If $v \in \Phi \smallsetminus \partial\Phi$ is the initial point of an edge whose endpoint v' lies on the arc joining x_{i-1} to x_i in $\partial\Phi$, then γ_v is a distance less than $\rho/8 + 2\rho_k + \rho/2$ from either $w_{i-1} \cdot p$ or $w_i \cdot p$, depending on which side of the midpoint of the arc v' lies (where "midpoint" is measured in the arc length pulled back from $\tilde{M}$).

For each $i = 1, \ldots, n$ we collapse all but one of the edges along the arc of $\partial\Phi$ joining x_{i-1} to x_i; the edge containing the midpoint is not collapsed[23], and its image in the quotient disc $\overline{\Phi}$ is labelled with the i-th letter of w. The image in $\overline{\Phi}$ of the quotient of the edge $[v, v']$ discussed in the previous paragraph is labelled either $\gamma_v^{-1} w_{i-1}$ or $\gamma_v^{-1} w_i$, according to the side of the midpoint on which v' lies. (This label will be an element of $\mathcal{A}$ because $\rho/8 + 2\rho_k + \rho/2 < \rho$.)

At this stage we have constructed a combinatorial 2-disc $\overline{\Phi}$ with a label from $\mathcal{A}$ on each directed 1-cell. The label on the boundary circle $\partial\overline{\Phi}$ is our original null-homotopic word w. The label on the boundary cycle of each 2-cell is, by construction, a word of length at most 12 in the letters $\mathcal{A}$ that represents the identity in Γ, because the faces of Φ, and hence $\overline{\Phi}$, are k-gons with $k \leqslant 12$. Thus $\overline{\Phi}$ is a van Kampen diagram for w over our chosen presentation of $\Gamma = \pi_1 M$.

The Cellulation Lemma gave us Φ and told us that it had at most $8N$ faces, where $N = \lambda_k(\mathrm{Fill}_0^M(\rho\,|w|) + \rho\,|w| + 1)$. And $\overline{\Phi}$ has the same number of faces as Φ. Thus we have established the desired upper bound on the algebraic area of the arbitrary null-homotopic word w, and we deduce that $\delta_\Gamma \preceq \mathrm{Fill}_0^M$. $\square$

6 Linear and Quadratic Dehn Functions

In this section we shall see that the groups that have linear Dehn functions are precisely those that are negatively curved on the large scale, i.e. *hyperbolic* in

[23] this involves a choice if the midpoint is a vertex

the sense of 6.1.3. This fundamental insight is due to Misha Gromov [55].

We shall also discuss the weaker link between non-positive curvature and the class of groups that have a quadratic Dehn function.

6.1 Hyperbolicity: from Dehn to Gromov

Given a finite set of generators $\mathcal{A}$ for a group Γ, one would have a particularly efficient algorithm for solving the word problem if one could construct a finite list of words $u_1, v_1, u_2, v_2, \dots, u_n, v_n$, with $u_i =_\Gamma v_i$ and $|v_i| < |u_i|$, such that every freely-reduced word in the letters $\mathcal{A}^{\pm 1}$ that represents $1 \in \Gamma$ contains at least one of the u_i as a subword.

If such a list of words exists then one proceeds as follows: given an arbitrary reduced word w, look for subwords of the form u_i; if there is no such subword, stop and declare that w does not represent $1 \in \Gamma$; if u_i occurs as a subword, replace u_i with v_i, freely reduce the resulting word w' and then repeat the search for subwords of the form u_j (noting that $w = w'$ in Γ). Proceeding in this way, after at most $|w|$ steps one will have either reduced w to the empty word (in which case $w = 1$ in Γ) or else verified that $w \neq 1$ in Γ.

Definition 6.1.1 *When it exists, the above procedure for solving the word problem is called a* Dehn algorithm *for Γ; it is encoded in $\langle \mathcal{A} \mid u_1 v_1^{-1}, \dots, u_n v_n^{-1} \rangle$, which we call a* Dehn presentation.

Max Dehn proved that Fuchsian groups admit Dehn presentations [35]. Jim Cannon proved that the fundamental groups of all closed negatively curved manifolds admit Dehn presentations [27]. The following *small cancellation* condition provides many other examples (see [66] Chapter V).

Example 6.1.2 Let $\langle \mathcal{A} \mid \mathcal{R} \rangle$ be a finite presentation in which each relator is freely reduced. Assume that if $r \in \mathcal{R}$ then r^{-1} and every cyclic permutation of r is in $\mathcal{R}$. And suppose that whenever there exist distinct $r, r' \in \mathcal{R}$ with a common prefix u (i.e. $r \equiv uv$ and $r' \equiv uv'$), the inequality $|u| < |r|/6$ holds. Then $\langle \mathcal{A} \mid \mathcal{R} \rangle$ is a Dehn presentation.

It requires only a moment's clear thought to see that the existence of a Dehn algorithm for a group Γ implies that Γ has a linear Dehn function (cf. paragraph 1.2). A more profound observation is that the converse is also true. The proof of this fact is indirect, proceeding via Gromov's notion of a hyperbolic group [55].

Gromov made the following remarkable discovery: the simple geometric condition given in (6.1.3) forces a geodesic metric space, regardless of its local structure, to exhibit many of the large-scale features that one associates with simply-connected manifolds of negative curvature. Thus he was able to extend the power of negative curvature well beyond its traditional realm[24] in Riemannian geometry. This stripping away of extraneous structure leads to a deeper understanding

[24] The work of H. Busemann and, more particularly, A.D. Alexandrov, had already expanded the range of spaces in which one can discuss negative and non-positive curvature (see [22]), but that work was based on local definitions of curvature, whereas in Gromov's approach one ignores the local structure of the space.

of the fundamental groups of closed negatively curved manifolds, and extends such an understanding to much wider classes of groups.

Definition 6.1.3 *A geodesic metric space X is* hyperbolic *(in the sense of Gromov) if there exists a constant $\eta > 0$ such that for every geodesic triangle*[25] *$\Delta \subseteq X$, each edge of Δ lies in the η-neighbourhood of the union of the other two edges. (One writes "X* is η-hyperbolic*" when it is useful to specify the constant.)*

A finitely generated group Γ is said to be *hyperbolic* if its Cayley graph[26] is η-hyperbolic for some $\eta > 0$.

Exercises 6.1.4 (i) Prove that real hyperbolic space $\mathbb{H}^n$ is hyperbolic in the above sense and find the optimal η. (Hint: There is a bound on the area of semicircular discs that can be inscribed in geodesic triangles in $\mathbb{H}^2$.)

(ii) Deduce that the universal covering X of any closed manifold of negative sectional curvature is hyperbolic in the sense of Gromov. (Hint: If one scales the metric so that the curvature of X is bounded above by -1, then every geodesic triangle $\Delta \subseteq X$ is the image of a non-expanding map $\phi : \overline{\Delta} \to \Delta$, where $\overline{\Delta}$ is a triangle in $\mathbb{H}^2$ and the restriction of ϕ to each edge of $\overline{\Delta}$ is an isometry. This is called the CAT(-1) inequality [22].)

The following results are due to Gromov [55] (see also Cannon [28]). Detailed references and proofs can be found in Chapter III.Γ of [22].

Theorem 6.1.5 *The following statements are equivalent for finitely presented groups Γ:*

(1) *Γ is a hyperbolic group.*

(2) *Γ has a finite Dehn presentation.*

(3) *Γ has a linear Dehn function.*

(4) *The Dehn function of Γ is sub-quadratic (i.e. $\delta_\Gamma(n) = o(n^2)$).*

Proceeding in cyclic order, the only non-trivial implications are (4) $\Rightarrow$ (1) and (1) $\Rightarrow$ (2). We shall not discuss (4) $\Rightarrow$ (1) except to say that Cornelia Druţu [38] recently discovered an elegant proof that uses asymptotic cones (cf. 3.1.6).

The proof that (1) $\Rightarrow$ (2) requires an understanding of the following types of locally-efficient paths. Let $I \subset \mathbb{R}$ be an interval and let X be a metric space. A map $c : I \to X$ is called a k*-local* geodesic if $d(c(t), c(t')) = |t - t'|$ for all $t, t' \in I$ with $|t - t'| \leqslant k$. And c is called a (λ, ε)*-quasi-geodesic* if

$$\frac{1}{\lambda}\,|t - t'| - \varepsilon \leqslant d(c(t), c(t')) \leqslant \lambda\,|t - t'| + \varepsilon$$

[25] See Appendix B for definitions such as that of a triangle in an arbitrary metric space.

[26] The ambiguity that arises from the fact that we have not specified a generating set is removed by Exercise 6.1.9(2).

for all $t, t' \in I$.

In hyperbolic spaces one has the following local criterion for recognising certain quasi-geodesics (see [22] page 405).

Lemma 6.1.6 *If X is η-hyperbolic then every 8η-local geodesic in X is a (λ, ε)-quasi-geodesic, where the constant $\lambda > 0$ depends only on η, and ε is less than 8η.*

The implication (1) $\Rightarrow$ (2) in Theorem 6.1.5 follows easily from this lemma:

Exercise 6.1.7 Suppose that the Cayley graph of Γ with respect to the finite generating set $\mathcal{A}$ is η-hyperbolic. Let $\mathcal{R}$ be the set of words $u_i v_i^{-1}$, where u_i runs over all words of length $\leqslant 8\eta$ in the letters $\mathcal{A}^{\pm 1}$ for which there exists a word v_i with $|v_i| < |u_i|$ and $u_i = v_i$ in Γ. Show that $\langle \mathcal{A} \mid \mathcal{R} \rangle$ is a Dehn presentation.

The following stability property of quasi-geodesics marks an important difference between spaces of non-positive curvature and spaces of strictly negative curvature (see [22] page 401).

Proposition 6.1.8 *For all $\eta, \lambda, \varepsilon > 0$ there exists $R(\eta, \lambda, \varepsilon) > 0$ such that: if X is η-hyperbolic and $c : [a, b] \to X$ is (λ, ε)-quasi-geodesic with endpoints p and q, then the Hausdorff distance between the image of c and each geodesic segment joining p to q is less than $R(\eta, \lambda, \varepsilon)$.*

This proposition provides a proof (independent of the Filling Theorem) that the fundamental groups of closed negatively curved manifolds have linear Dehn functions – see 6.1.4(ii) and 6.1.9(iii).

The following exercises require the reader to understand certain items from Appendix A, namely the definition of quasi-isometry, the Švarc-Milnor Lemma and A.1.3(ii).

Exercise 6.1.9 (i) Let X be a geodesic space. If X is quasi-isometric to a η-hyperbolic space, then X is η'-hyperbolic for some $\eta' > 0$. (Hint: Consider quasi-geodesic triangles.)

(ii) If the Cayley graph of a group with respect to one finite generating set is hyperbolic, then so is the Cayley graph of that group with respect to any other finite generating set.

(iii) If a group acts properly and cocompactly by isometries on a hyperbolic geodesic space, then that group has a linear Dehn function.

We refer the reader to Chapter III.Γ of [22] for an introduction to the rich theory of hyperbolic metric spaces (the references given therein will also point the reader to recent developments in this active field). Here are a few of the basic properties of hyperbolic groups.

Theorem 6.1.10 *If a group Γ has a linear Dehn function then:*

(1) Γ *does not contain* $\mathbb{Z}^2$ *;*

(2) Γ *has a solvable conjugacy problem;*

(3) Γ *has only finitely many conjugacy classes of finite subgroups;*

(4) Γ *acts on a contractible simplicial complex with compact quotient and finite stabilizers.*

(5) *Let* $\mathcal{A}$ *be a finite generating set for* Γ *and let* d *be the associated word metric. Define* $\tau(\gamma) = \lim_{n\to\infty} d(1, \gamma^n)/n$. *Then there is an integer* N *such that* $\{N\,\tau(\gamma) \mid \gamma \in \Gamma \smallsetminus \{1\}\}$ *is a set of positive integers.*

6.2 Quadratic Dehn Functions and Non-Positive Curvature

If a geodesic metric space X is complete, 1-connected and non-positively curved in the sense of A.D. Alexandrov (see Appendix B), then its metric is *convex* in the sense that $d(c(t), c'(t)) \leqslant t\, d(c(1), c'(1)) + (1-t)\, d(c(0), c'(0))$ for all geodesics $c, c' : [0,1] \to X$ parameterized by arc length. This class of spaces includes the universal covering $\tilde{M}$ of any compact Riemannian manifold whose sectional curvatures are non-positive, and hence the following result applies to the fundamental groups of such manifolds (acting by deck transformation on $\tilde{M}$). It also applies to cocompact lattices in semisimple Lie groups (cf. 3.3.7).

Theorem 6.2.1 *Let* X *be a complete geodesic space whose metric is convex. If the group* Γ *acts properly by isometries on* X *and the quotient of this action is compact, then* Γ *is finitely presented and its Dehn function is either linear or quadratic.*

The following proof is adapted from [5] and [22], and has earlier origins, e.g. [42].

Proof The point of the proof is to construct the diagram shown in figure 6.2.3. Let d be the metric on X. Fix $p \in X$ and let $\rho \geqslant 1$ be such that the balls of radius ρ about the Γ-orbit of p cover X. Let c_γ be the arc-length parameterization of the unique geodesic segment joining p to $\gamma.p$. Let $\mathcal{A} \subset \Gamma$ be the set of $\gamma \in \Gamma$ such that $d(p, \gamma.p) \leqslant 3\rho$. Given $\gamma \in \Gamma$, let m be the least integer greater than $d(p, \gamma.p)/\rho$ and for each positive integer $t < m$ choose $\gamma_t \in \Gamma$ with $d(c_\gamma(\rho t), \gamma_t.p) \leqslant \rho$. Define $\gamma_0 = 1$ and $\gamma_m = \gamma$.

Consider the word $\sigma_\gamma := a_1 \dots a_m$ where $a_i := \gamma_{i-1}^{-1}\gamma_i \in \mathcal{A}$ for $i = 1, \dots, m$. With an eye on future generalisations, we write $\sigma_\gamma(i)$ instead of γ_i to denote the image in Γ of the i-th prefix of σ_γ; by definition $\sigma_\gamma(i) = \gamma$ if $i > m$. (In general we write $w(i)$ for the image in Γ of the i-th prefix of any word w.)

It follows from the convexity of the metric on X that in the word metric $d_{\mathcal{A}}$ on Γ one has

$$d_{\mathcal{A}}(\sigma_\gamma(i), \sigma_{\gamma'}(i)) \leqslant 3\, d_{\mathcal{A}}(\gamma, \gamma') \tag{6.2.2}$$

for all $\gamma, \gamma' \in \Gamma$ and all integers $i > 0$ (see Exercise 6.2.4(i)). We shall use this inequality to construct efficient diagrams for null-homotopic words.

Let w be a null-homotopic word, of length n say. We draw an oriented circle in $\mathbb{R}^2$, mark vertices $v_0, \dots, v_{n-1}$ (in cyclic order) on the circle and label the oriented arc (v_{i-1}, v_i) with the i-th letter of w (indices mod n). We then connect v_0 to each of the vertices v_i with a line segment $[v_0, v_i]$ divided into $|\sigma_{w(i)}|$ 1-cells; these 1-cells are oriented and labelled by the letters of $\sigma_{w(i)}$ in the obvious manner. Define $\sigma_{w(0)} = \sigma_{w(n)}$ to be the empty word, and for $j \geqslant |\sigma_{w(i)}|$ define "the j-th vertex of $[v_0, v_i]$" to be v_i. Let $J(i) = \max\{|\sigma_{w(i)}|, |\sigma_{w(i+1)}|\}$.

We complete the construction of our diagram for w by introducing an edge from the j-th vertex of $[v_0, v_i]$ to the j-th vertex of $[v_0, v_{i+1}]$ for $i = 0, \dots, n-1$ and $j = 1, \dots, J(i)$; this edge is labelled by a word of minimal length that equals $\sigma_{w(i)}(j)^{-1}\sigma_{w(i+1)}(j) \in \Gamma$; according to (6.2.2) this word has length at most 3.

We have constructed a diagram over $\mathcal{A}$ with boundary label w, where w is an arbitrary null-homotopic word. The face labels are null-homotopic words of length $\leqslant 8$; let $\mathcal{R}$ be the set of all such words. Lemma 4.1.2 tells us that $\Gamma = \langle \mathcal{A} \mid \mathcal{R} \rangle$ and that $\mathrm{Area}_a(w)$ is at most the number of faces in the diagram. Thus $\mathrm{Area}_a(w) \leqslant |w| \max\{|\sigma_{w(i)}| : i \leqslant |w|\}$. And since $d_{\mathcal{A}}(1, w(i)) \leqslant |w|/2$ for all i, Exercise 6.2.4(ii) tells us that $\mathrm{Area}_a(w) \leqslant (3/2)|w|^2$. $\square$

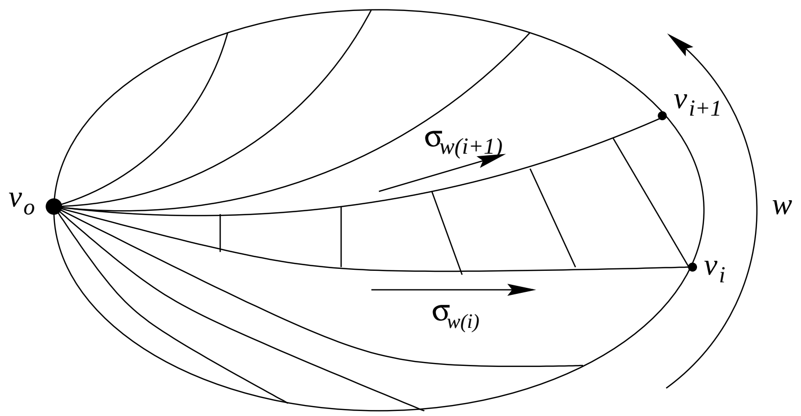

Figure 6.2.3 *Using the combing σ_γ to construct a van Kampen diagram*

Exercises 6.2.4 (i) Establish the inequality 6.2.2. (Hint: If $m = d_{\mathcal{A}}(\gamma, \gamma')$ then $d(\gamma.p, \gamma'.p) \leqslant 3m\rho$. Hence, by the convexity of the metric, $d(c_\gamma(\rho t), c_{\gamma'}(\rho t) \leqslant 3m\rho$ for all $t > 0$. Recall that, by definition, $\sigma_g(t) = g_t$. Divide the geodesic $[c_\gamma(\rho t), c_{\gamma'}(\rho t)]$ into $3m$ segments of equal length, and associate to each division point a closest point of $\Gamma.p$, with γ_t and γ'_t associated to the endpoints.)

(ii) Deduce that for all $\gamma \in \Gamma$ the length of the word σ_γ in the above proof is at most $3\, d_{\mathcal{A}}(1, \gamma)$.

6.3 Automatic Groups

The intensive study of isoperimetric inequalities for finitely presented groups began in the late 1980s. It emerged primarily from the work of Gromov [55], but a certain impetus also came from the theory of automatic groups. This theory sprang from conversations between Jim Cannon and Bill Thurston and grew into a rich theory due to a team effort orchestrated by David Epstein – see [42].

Roughly speaking, a group Γ with finite generating set $\mathcal{A}$ is *automatic* if one can construct its Cayley graph by computations on finite state automata: there must exist a set of words $\mathcal{L} = \{\sigma_\gamma \mid \gamma \in \Gamma\}$ in the letters $\mathcal{A}^{\pm 1}$, with $\sigma_\gamma = \gamma$ in Γ, such that membership of $\mathcal{L}$ can be determined by a finite state automaton (FSA); and for each $a \in \mathcal{A}$ there must exist a FSA that recognises those pairs of words $(\sigma_\gamma, \sigma_{\gamma'})$ for which $\gamma' = \gamma a$.

The finiteness of these FSA forces the existence of constants $k, K > 0$ such that $|\sigma_\gamma| \leqslant k\, d_{\mathcal{A}}(1, \gamma)$ and

$$d_{\mathcal{A}}(\sigma_\gamma(i), \sigma_{\gamma'}(i)) \leqslant K\, d_{\mathcal{A}}(\gamma, \gamma') \tag{6.3.1}$$

for all $\gamma, \gamma' \in \Gamma$. By using the normal form $\mathcal{L}$ in place of the words σ_γ constructed in the proof of (6.2.1) we obtain:

Theorem 6.3.2 *If Γ is automatic then it is finitely presented and its Dehn function is linear or quadratic.*

Automatic groups form a large class. This class includes many groups that do not arise in the setting of Theorem 6.2.1, for example central extensions of hyperbolic groups [77].

In Chapter 9 of [42] Epstein and Thurston determine which geometrizable 3-manifolds have automatic fundamental groups, and Theorem 3.3.1 follows from this work. All mapping class groups are automatic [76].

6.4 The Link with Non-Positive Curvature is Limited

In analogy with the theory of hyperbolic groups, one can develop a theory of *semihyperbolic groups*, defined by a coarse geometric constraint that forces such groups to satisfy most of the useful properties enjoyed by the fundamental groups of compact non-positively curved manifolds (cf. Alonso and Bridson [5] and Gromov [56]).

With Theorems 3.1.6 and 6.1.10 in mind, one might hope that requiring a group to satisfy a quadratic isoperimetric inequality would force it to behave in a "semihyperbolic" manner, satisfying a list of properties analogous to (6.1.10). The examples that we have seen thus far support this hope to some extent — abelian groups, hyperbolic groups, automatic groups, fundamental groups of compact non-positively curved spaces, $\mathrm{SL}(n, \mathbb{Z})$ for $n \geqslant 4$, various nilpotent groups N, and those non-uniform lattices in rank 1 Lie groups that have these N as cusp groups. But the examples discovered more recently indicate that the class of groups that have a quadratic Dehn function is wilder than this list would suggest, but quite how wild is not clear. For example, it is unknown if a group Γ

that has a quadratic[27] Dehn function can have an unsolvable conjugacy problem (it is conjectured that if such a group exists, it should not have a 2-dimensional $K(\Gamma, 1)$).

Besides the property that defines them, the most significant property that is known to be enjoyed by groups with quadratic Dehn functions is the fact that their asymptotic cones are all simply-connected [88]. This property is not enjoyed by all groups with polynomial Dehn functions [18].

7 Techniques for Estimating Isoperimetric Functions

This section contains a sample of the methods that have been developed to calculate Dehn functions. The techniques that I shall describe have been used widely, but I must emphasise that this is only a sample, not a thorough survey. This sample is biased in favour of the methods that I have found most useful in my own work.

7.1 Upper Bounds

In general it is easier to obtain upper bounds on Dehn functions than it is to obtain lower bounds. Indeed whenever one has an explicit solution to the word problem in a finitely presented group, one can look for an upper bound on the Dehn function by analysing the use of relations in that solution (cf. paragraph 1.1). Thus there are many *direct methods* for obtaining upper bounds, each adapted to the groups at hand. We have already seen examples of such methods in (3.1.2), (2.2), (6.1.1), and (3.2.3). One might also think of results such as 3.4.1 in this light. Direct methods of a geometric nature are to be found in many of the papers listed in the bibliography, e.g. [19], [15], [95] and [18].

The following general method for obtaining upper bounds on Dehn functions has been used in many contexts.

Using Combings to Get Upper Bounds. Let Γ be a group with finite generating set $\mathcal{A}$ and let d be the associated word metric. A *combing* (normal form) for Γ is a set of words $\{\sigma_\gamma \mid \gamma \in \Gamma\}$ in the letters $\mathcal{A}^{\pm 1}$ such that $\sigma_\gamma = \gamma$ in Γ. Whenever one can find a geometrically-efficient combing for a group Γ one can estimate the Dehn function δ_Γ by modifying the proof of Theorem 6.2.1. The control that one needs in order to get non-trivial bounds is remarkably weak [16]. We content ourselves with one of the simplest and most widely used methods of control, wherein one weakens *the fellow-traveller property* (6.3.1) by allowing reparameterizations of the words σ_γ (thought of as paths in the Cayley graph of Γ).

Definition 7.1.1 *Let*

$$R = \{\rho : \mathbb{N} \to \mathbb{N} \mid \rho(0) = 0;\ \rho(n+1) \in \{\rho(n), \rho(n)+1\}\ \forall n;\ \rho \textit{ unbounded}\}.$$

[27] There do exist examples with cubic Dehn functions, [20] Example 2.9.

Given words w_1, w_2 in the letters $\mathcal{A}^{\pm 1}$, define

$$D(w_1, w_2) \;=\; \min_{\rho,\rho' \in R} \big\{ \max_{t \in \mathbb{N}} \{d(w_1(\rho(t)), w_2(\rho'(t))\}\big\}.$$

A combing $\gamma \mapsto \sigma_\gamma$ is said to satisfy the asynchronous fellow-traveller property if there is a constant $K > 0$ such that

$$D(\sigma_\gamma, \sigma_{\gamma'}) \leqslant K\, d_{\mathcal{A}}(\gamma, \gamma')$$

for all $\gamma, \gamma' \in \Gamma$. The length *of σ is a function $\mathbb{N} \to \mathbb{N}$:*

$$L_\sigma(n) := \max\{|\sigma_\gamma| \mid d_{\mathcal{A}}(1, \gamma) \leq n\}.$$

Proposition 7.1.2 *If a finitely generated group Γ admits a combing σ that satisfies the asynchronous fellow-traveller property, then Γ is finitely presented and its Dehn function satisfies $\delta_\Gamma(n) \preceq n L_\sigma(n)$. And regardless of the length of the combing, $\delta_\Gamma(n) \preceq 2^n$.*

Exercise 7.1.3 Prove the assertions in the first sentence of the above proposition. (Hint: Follow the construction of Figure 5 in the proof of Theorem 6.2.1, but instead of connecting $\sigma_{w(i)}(j)$ to $\sigma_{w(i+1)}(j)$ with a 1-cell, connect $\sigma_{w(i)}(\rho(j))$ to $\sigma_{w(i+1)}(\rho'(j))$, where ρ and ρ' are reparameterizations as in the definition of the asynchronous fellow-traveller property.)

Examples 7.1.4 (i) The upper bound described in Theorem 3.1.4 was established in [23] using the combings constructed in [17]. Given $\Gamma = \mathbb{Z}^m \rtimes_\phi \langle t \rangle$ one can write each $\gamma \in \Gamma$ uniquely in the form $t^n x$ with $x \in \mathbb{Z}^m$. One fixes a basis for $\mathbb{Z}^m$ and represents x by a word l_x that (viewed as a path in the lattice $\mathbb{Z}^m$) stays closest to the Euclidean segment $[0, x]$ in $\mathbb{R}^m = \mathbb{Z}^m \otimes \mathbb{R}$. One then defines $\sigma_\gamma = t^n l_x$, checks that σ satisfies the asynchronous fellow-traveller property and calculates that $L_\sigma(n) \simeq n\,\|\phi^n\|$ (see [23] page 215).

(ii) I proved in [17] that if a compact 3-manifold M satisfies the geometrization conjecture, then $\pi_1 M$ admits a combing that satisfies the asynchronous fellow-traveller property, whence the exponential upper bound in Theorem 3.3.1.

7.2 Lower Bounds

t-corridors and t-rings. *t-corridors* and *t-rings* are particular types of subdiagrams that one gets in van Kampen diagrams over presentations $\langle \mathcal{A}, t \mid \mathbb{R} \rangle$ where the group presented retracts onto $\langle t \rangle$. We refer to [21] for a careful treatment, but point out that although this is where *t-corridors* were named and systematised, they were in use much earlier, e.g. in Rips's geometric proof of the unsolvability of the word problem (see the inside cover of [94]).

Consider the presentation $\Gamma = \langle \mathcal{A}, t_1, \dots, t_n \mid \mathcal{R}\rangle$, where the symbols t_j are not elements of $\mathcal{A}$ and the only relators involving any t_j are of the form $t_j u_i t_j^{-1} v_i \in \mathcal{R}$, where $u_i, v_i \in F(\mathcal{A})$. Consider a van Kampen diagram D over such a presentation and focus on an edge ε in the boundary labelled $t \in \{t_1, \dots, t_n\}$. If this edge lies in the boundary of a 2-cell, then the boundary cycle of this 2-cell (read with suitable orientation from ε) has the form $tut^{-1}v$ with $u, v \in F(\mathcal{A})$. In particular, there is a unique edge other than ε in the boundary of the 2-cell that is labelled t; crossing this edge we enter another 2-cell with a similar boundary label; by iterating the argument we get a chain of 2-cells running across the diagram; this chain terminates at an edge of ∂D which (following the orientation of ∂D in the direction of our original edge ε) is labelled t^{-1}. This chain of 2-cells is called a *t-corridor*.

Topologically, a t-corridor is a map $[0,1] \times [0,1] \to D$ that is injective on $[0,1] \times (0,1)$. We make this map a morphism of labelled combinatorial 2-complexes by pulling back the cell structure and labelling from D. The labels on the 1-cells in $[0,1] \times \{0,1\}$ (the top and bottom of the corridor) are letters from $\mathcal{A}^{\pm 1}$; the remaining 1-cells are of the form $\{s\} \times [0,1]$, and these are labelled t.

A *t-ring* is defined similarly: it consists of a chain of 2-cells giving a combinatorial map $\phi : \mathbb{S}^1 \times [0,1] \to D$ that is injective on $\mathbb{S}^1 \times (0,1)$; in $\mathbb{S}^1 \times [0,1]$ the 1-cells of the form $\{\theta\} \times [0,1]$ are labelled t; the remaining 1-cells are contained in $\mathbb{S}^1 \times \{0,1\}$ and are labelled by letters from $\mathcal{A}^{\pm 1}$; the map ϕ is label-preserving.

Much of the utility of t-corridors and t-rings rests on the following observations:

Exercise 7.2.1 Let t_i, u_i, v_i, Γ and D by as in the preceding discussion. Prove:

(i) Distinct t-corridors and t-rings have disjoint interiors.

(ii) If P is the edge-path in D running along the top or bottom of a t-corridor, then P is labelled by a word in the letters $\mathcal{A}^{\pm 1}$ that is equal in Γ to the words labelling the subarcs of ∂D which share the endpoints of P (given appropriate orientations),

(iii) and if $k = \min\{\max_i |u_i|, \max_i |v_i|\}$, then the number of 2-cells in the t-corridor is at least $1/k$ times the length of P.

(iv) The words labelling the inner and outer boundary cycles of a t-ring are null-homotopic.

(v) If D contains a 2-cell that has an edge labelled t in its boundary, then D contains either a t-corridor or a t-ring.

Instead of indulging in a general discussion, let me give one proposition to illustrate the utility of t-corridors and one to illustrate the utility of t-rings.

Proposition 7.2.2 *Let ϕ be an automorphism of the finitely presented group $B = \langle \mathcal{A} \mid \mathcal{S}\rangle$. For each $a \in \mathcal{A}$, choose a word $v_a \in F(\mathcal{A})$ representing $\phi(a) \in B$. Let $\mathcal{R} = \mathcal{S} \cup \{t_j^{-1} a t_j = v_a \mid a \in \mathcal{A}, j = 1, 2\}$ and define $\Gamma := \langle \mathcal{A}, t_1, t_2 \mid \mathcal{R}\rangle$. Then the Dehn function of Γ is $\simeq$ bounded below by*

$$n \mapsto n \max_b \{d_{\mathcal{A}}(1, \phi^n(b)) \mid d_{\mathcal{A}}(1, b) \leqslant n\}.$$

Proof For each positive integer n, we choose a word β of length at most n in the generators $\mathcal{A}^{\pm 1}$ so as to maximize $d_{\mathcal{A}}(1, \phi^n(b))$, where b is the image of β in Γ. Let $u_n := t_1^{-n}\beta t_1^n$ and let $w_n := u_n(t_2t_1^{-1})^n u_n^{-1}(t_2t_1^{-1})^{-n}$, a word of length at most $10n$. Note that $w_n = 1$ in Γ. Note also that no proper subword of w_n is equal to $1 \in \Gamma$ (one sees this easily using the natural retraction $\Gamma \to F(t_1, t_2)$ and the fact that $t_j^{-i}\beta t_j^i \neq 1$ for all i). It follows that any van Kampen diagram for w_n is a disc, in particular every edge of ∂D lies in the closure of some 2-cell, and therefore a t_j-corridor emanates from each edge of ∂D labelled t_j, for $j = 1, 2$.

The simple fact that distinct t_j-corridors cannot cross (fact 7.2.1(i)) implies that the pattern of t_2-corridors in any van Kampen diagram for w_n must be as shown in figure 7.2.3. The words in the letters $\mathcal{A}^{\pm 1}$ labelling the bottom of each of each t_2-corridor is equal in Γ to u_n. Hence (fact 7.2.1(iii)) each of these corridors contains at least $\frac{1}{k}\, d_{\mathcal{A}}(1, \phi^n(b))$ 2-cells, where k is the length of the longest of the words v_a. And there are n such corridors. $\square$

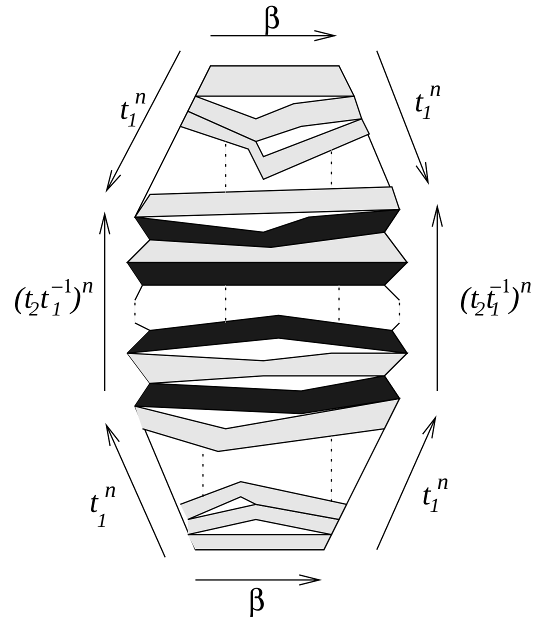

Figure 7.2.3 *The pattern of t_j-corridors*

Exercises 7.2.4 (i) Let $\phi \in \mathrm{GL}(n,\mathbb{Z})$ be a unipotent matrix and let $\Gamma = \mathbb{Z}^m \rtimes F(t_1,t_2)$, where the generators t_1 and t_2 of the free group $F(t_1,t_2)$ both act on $\mathbb{Z}^m$ as ϕ. Deduce from the above proposition and your proof of (3.1.5) that the Dehn function of Γ is bounded below by a polynomial of degree $c+1$, where c is the size of the largest elementary block in the Jordan form of ϕ. Adapt 7.1.4(i) to deduce that in fact $\delta_\Gamma(n) \simeq n^{c+1}$. (If you get stuck, refer to [18].)

(ii) Britton's Lemma states that, given an HNN extension $G*_\phi = (G, t \mid t^{-1}st = \phi(s), \forall s \in S)$, and a generating set $\mathcal{A}$ for G, every null-homotopic word in the letters $(\mathcal{A} \cup \{t\})^{\pm 1}$ either contains no occurrences of $t^{\pm 1}$, or else contains a subword $t^\varepsilon u t^{-\varepsilon}$, where $\varepsilon = \pm 1$ and u is a word in the letters $\mathcal{A}^{\pm 1}$ that lies in $\langle S \rangle$ if $\varepsilon = -1$ and lies in $\langle \phi(S) \rangle$ if $\varepsilon = 1$.

Use t-corridors to prove Britton's Lemma.

Proposition 7.2.5 *Let G be a finitely presented group, let $L, L' \subset G$ be finitely generated subgroups that are free, let $\phi : L \to L'$ be an isomorphism, and let $\Gamma = G*_\phi$ be the associated HNN extension. Then $\delta_\Gamma \succeq \delta_G$.*

Exercises 7.2.6 (i) Let D be a van Kampen diagram for a null-homotopic word w over a presentation $\langle \mathcal{A} \mid \mathcal{R} \rangle$, and let u be the label on a simple closed loop c in the 1-skeleton of D. Prove that if $\mathrm{Area}_c\, D = \mathrm{Area}_a w$, then the number of 2-cells in the sub-diagram enclosed by c is $\mathrm{Area}_a u$. Extend this result to non-crossing loops[28].

(ii) Prove Proposition 7.2.5. (Hint: $\Gamma = \langle \mathcal{A}, t \mid \mathcal{R}, t^{-1}lt = \phi(l), l \in S \rangle$ where $G = \langle \mathcal{A} \mid \mathcal{R} \rangle$ and $S \subset \mathcal{A}$ is a basis for L. Given a word $w \in F(\mathcal{A})$ with $w = 1$ in Γ, take a van Kampen diagram D with $\mathrm{Area}_c\, D = \mathrm{Area}_a w$. Use (i) and the fact that L is free to argue that D contains no t-rings and hence is a diagram over $\langle \mathcal{A} \mid \mathcal{R} \rangle$.)

Cohomological Methods. Both Gersten and Gromov have developed cohomological methods for obtaining lower bounds on Dehn functions. In particular, Gersten [48] developed an ℓ_∞-cohomology theory which, among other things, allows one to recover results obtained using t-corridors in a more elegant and systematic manner. It would take too long to explain these ideas here, so we refer the reader to [48]. We content ourselves with a more simple-minded result that uses de Rham cohomology. (We give this result in part because it resonates with ideas in Section 5).

The statement of the following lemma is phrased in the vocabulary introduced in (4.3). The constant A_ω is defined to be the maximum of the integrals $\int_E \omega$ where E is a 2-cell mapped into $\tilde{M}$ by $\tilde{K}(\mathcal{A} : \mathcal{R}) \to \tilde{M}$.

Lemma 7.2.7 *Let M be a smooth, closed Riemannian manifold with fundamental group $\Gamma = \langle \mathcal{A} \mid \mathcal{R} \rangle$ and let ω be a Γ-invariant closed 2-form on $\tilde{M}$. If D*

[28] A *non-crossing loop* is the restriction to $S^1 \times \{1\}$ of a map $S^1 \times [0,1] \to \mathbb{R}^2$ that is injective on $S^1 \times [0,1)$.

is a van Kampen diagram with boundary label w, and h_D is the map described in (4.3), then

$$\int_D h_D^* \omega \leqslant A_\omega \ \mathrm{Area}_a(w).$$

Proof The integral $\int_D h_D^*\omega$ is well-defined because h_D is differentiable except on a set of measure zero. If D' is a second van Kampen diagram for w, then one can regard $-h_D \cup h_{D'}$ as a 2-cycle in $\tilde{M}$, and hence $\int_D h_D^*\omega = \int_{D'} h_{D'}^*\omega$. And when $\mathrm{Area}_c\, D = \mathrm{Area}_a(w)$ the inequality is clear. □

The utility of this lemma stems from the fact that one does not need to understand the nature of least-area van Kampen diagrams in order to get a lower bound on their area: if one can locate *any* van Kampen diagram D for a given word w, then one gets a lower bound on $\mathrm{Area}_a(w)$ by integrating $h_D^*\omega$ over D. Moreover, by Stokes Theorem, if the 2-form ω is exact, say $\omega = \mathrm{d}\alpha$, then one can simply calculate $\int_{\hat{w}} \alpha$, thus avoiding the construction of diagrams altogether.

Example 7.2.8 In the case where $\phi \in \mathrm{Sp}(m, \mathbb{Z})$, Bridson and Pittet [23] established the lower bound in Theorem 3.1.4 by applying Lemma 7.2.7 to the standard symplectic form on $\mathbb{R}^m$.

Exploiting Asphericity. A group presentation $\langle \mathcal{A} \mid \mathcal{R} \rangle$ is called *aspherical* if the associated 2-complex $K(\mathcal{A}; \mathcal{R})$ is aspherical (i.e. its universal covering is contractible). One of the great joys of working with aspherical presentations is that when one finds an *embedded* van Kampen diagram one knows that it is of minimal area:

Lemma 7.2.9 *Suppose that $X = K(\mathcal{A}; \mathcal{R})$ is aspherical. Let D be a van Kampen diagram for w. If the associated map $D \to \widetilde{X}$ is injective on the complement of the 1-skeleton $D^{(1)}$, then the number of 2-cells in D is* $\mathrm{Area}_a(w)$.

Proof Let D' be a second van Kampen diagram for w. One can regard $D \cup D'$ as a 2-cycle in the cellular chain complex of $\tilde{X}$. Since there are no 3-cells and $H_2\tilde{X}$ is trivial (by Hurewicz), this 2-cycle must be zero. And since the 2-cells in the image of D are all distinct, each must cancel with some 2-cell in D'. Hence $\mathrm{Area}_c D \leqslant \mathrm{Area}_c D'$. And since D' was arbitrary, $\mathrm{Area}_a(w) = \mathrm{Area}_c\, D$. □

Examples 7.2.10 (i) $\mathbb{Z}^2 = \langle a, b \mid [a, b] \rangle$ is aspherical. Hence the area of the obvious (square) diagram for $w_n = a^{-n}b^{-n}a^n b^n$ equals $\mathrm{Area}_a(w_n)$ (cf. 3.1.2).

(ii) The presentation of B_m described in (3.2.1) is aspherical. Explicit disc diagrams show that $\delta_{B_m}(n) \succeq \varepsilon_m(n)$ — see Exercise 7.2.11.

(iii) A celebrated theorem of Roger Lyndon shows that 1-relator presentations are aspherical if the relation is not a proper power [66].

(iv) The natural presentations of free-by-free groups are aspherical and provide interesting examples of Dehn functions [20].

Exercises 7.2.11 (i) Let X denote the universal covering of the standard 2-complex of the presentation of B_m described in (3.2.1). The 1-skeleton $X^{(1)}$ is identified with the Cayley graph of B_m. Show that the loop in $X^{(1)}$ labelled $x_1^{-n}x_0x_1^nx_0^{-2^n}$ bounds an embedded disc $\Delta_n(x_0,x_1)$ that has (2^n-1) faces (2-cells). By juxtaposing two copies of $\Delta_n(x_0,x_1)$, construct a disc D_1 showing that $x_1^{-n}x_0x_1^nx_0x_1^{-n}x_0^{-1}x_1^nx_0^{-1}$ is a null-homotopic word of area $2(2^n-1)$.

(ii) Now suppose that $n=2^r$. By attaching four copies of a disc diagram $\Delta_r(x_1,x_2)$ to the segments of ∂D_1 labelled x_1^n, construct a disc diagram for $(x_2^{-r}x_1^{-1}x_2^r)x_0(x_2^{-r}x_1x_2^r)x_0(x_2^{-r}x_1^{-1}x_2^r)x_0^{-1}(x_2^{-r}x_1x_2^r)x_0^{-1}$ that has more than 2^{2^r} faces (2-cells).

Iterate this construction and use Lemma 7.2.9 to deduce that $\delta_{B_m}(n)\succeq\varepsilon_m(n)$.

Reprove this inequality using t-corridors instead of asphericity.

The following exercises lead the reader through the proof that the Dehn function of the group $G_{p,q}$ described in 3.1.10 is $\succeq n^{2\log_2 2p/q}$. If you get stuck during these exercises, refer to [15].

Exercises 7.2.12 (i) Let $f, g : [0, \infty) \to [0, \infty)$ be non-decreasing functions and let (n_i) be an increasing sequence of positive integers with $n_0 = 0$ and $n_{i+1} \leqslant Cn_i$ for all i, where $C > 0$ is constant. Show that if $f(n_i) \leqslant g(n_i)$ for all i, then $f \preccurlyeq g$. (Thus we see that to establish lower bounds on Dehn functions $\delta(n)$, it is only necessary to look at fairly sparse sequences of integers (n_i).)

(ii) Consider the presentation of $G_{p,q}$ given in (3.1.10). Prove that this presentation is aspherical. (Hint: One can build the 2-complex of the presentation as follows. Start with a torus corresponding to the subgroup $\mathrm{gp}\{a, b\} \subseteq G_{p,q}$ and fix a basepoint on it. Attach two cylinders (annuli) to the torus along simple curves through the basepoint – one end of each cylinder traces out a curve in the homotopy class a^q and the other ends trace out $a^p b^{\pm 1}$. The Seifert-van Kampen theorem shows that this complex has fundamental group $G_{p,q}$. The universal cover $\tilde{X}$ of this 2-complex is a contractible complex obtained by gluing planes indexed by the cosets of $\mathrm{gp}\{a, b\} \subset G_{p,q}$ along strips (copies of the line cross an interval) covering the annuli in the quotient.)

(iii) Complete the following outline to a proof that the Dehn function of $G_{p,q}$ is $\succeq n^\alpha$ where $\alpha = 2\log_2 2p/q$.

Let $w_0 = a^q$ and let $w_1 = sa^q s^{-1} t a^q t^{-1}$. Define words $w_k = sw_{k-1}a^{\epsilon_{k-1}}s^{-1}tw_{k-1}a^{\epsilon_{k-1}}t^{-1}$ with $0 \le \epsilon_{k-1} \le q-1$ so that $w_{k-1}a^{\epsilon_{k-1}}$ represents a power of a that is divisible by q. Show that $4(2^k) \le |w_k| \le (4q)2^k$ and that $w_k = a^{m_k}$ in $G_{p,q}$, where $m_k \geqslant q(2p/q)^k$.

Show that one can find embedded in $\tilde{X}$ a van Kampen diagram portraying the equality $w_k = a^{m_k}$. (See figure 7.2.13 – the large faces in this figure are diagrams over the sub-presentation $\langle a, b \mid [a, b] \rangle$.)

Let $W_k = [sw_{k-1}a^{\epsilon_{k-1}}s^{-1}, tw_{k-1}a^{\epsilon_{k-1}}t^{-1}]$. Show that W_k represents the identity in $G_{p,q}$ and describe a van Kampen diagram for W_k that embeds in $\tilde{X}$. Deduce that there is a constant $C > 0$ such that

$$\mathrm{Area}_a(W_k) \ \ge\ Cm_k^2 \ \ge\ Cq^2(2p/q)^{2k}.$$

Use (i) to conclude that the Dehn function of $G_{p,q}$ is bounded below by $n \mapsto n^{2\log_2 2p/q}$.

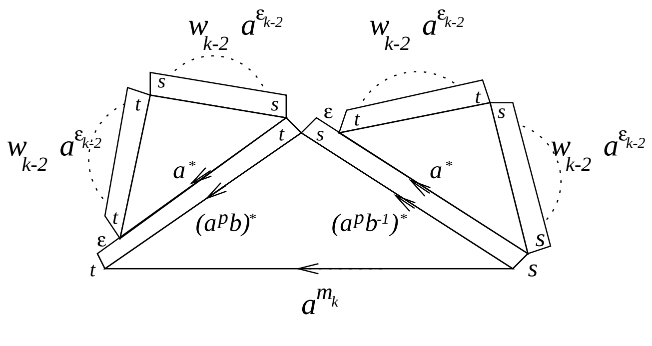

Figure 7.2.13 *The diagram portraying $w_k = a^{m_k}$*

Calculating in Abelian Quotients. Let $\Gamma = \langle \mathcal{A} \mid \mathcal{R} \rangle$ and let K be the subgroup of $F = F(\mathcal{A})$ generated by the set of elements $\mathcal{C} = \{x^{-1}rx \mid x \in F(\mathcal{A}),\ r \in \mathcal{R}\}$. By definition, $\text{Area}_a w$ is the least number N for which there is an equality $w = c_1 \dots c_N$ with $c_i \in \mathcal{C}^{\pm 1}$. One anticipates that the task of estimating N would be easier if one were working with sums in abelian groups rather products in free groups. With this in mind (and motivated by results of Gersten [45]) Baumslag, Miller and Short [10] look at the projection of equalities such as the one above into the abelianization of K, i.e. the *relation module*[29] of the presentation $\langle \mathcal{A} \mid \mathcal{R} \rangle$. They also consider what happens when one projects further, onto $K/[K, F]$.

Thus they define the *abelianized isoperimetric function* Φ_Γ^{ab} by analogy with the Dehn function (1.2.2), replacing $\text{Area}_a w$ by $\text{Area}_a^{\text{ab}} w$, which is defined to be the least integer N for which there is an equality

$$w = \sum_{i=1}^{N} x_i^{-1} r_i x_i$$

in $K/[K, K]$, with $r_i \in \mathcal{R}^{\pm 1}$ and $x_i \in F$. And they define the *centralized isoperimetric function* $\Phi_\Gamma^{\text{cent}}$ by counting the minimum number of summands required to express w in $K/[K, F]$. Baumslag *et al.* prove that each of these functions is $\simeq$ independent of the chosen finite presentation of Γ.

Note the obvious inequalities

$$\Phi_\Gamma^{\text{cent}} \preceq \Phi_\Gamma^{\text{ab}} \preceq \delta_\Gamma.$$

From the Hopf formula ([25] page 41) one sees that $K/[K, F]$ is a direct sum of a free abelian group and $H_2(\Gamma, \mathbb{Z})$, and there is a well-developed technology for calculating in $H_2(\Gamma, \mathbb{Z})$ – in particular one has Fox's free differential calculus. By using this calculus Baumslag *et al.* obtain bounds on $\Phi_\Gamma^{\text{cent}}$ for various groups. In certain cases they are also able to show that $\Phi_\Gamma^{\text{cent}} \simeq \delta_\Gamma$. In this way they were able to calculate the Dehn functions of free nilpotent groups, thus exemplifying the merits of the aphorism that the homological approach works best for groups that contain a lot of commutivity.

Exercise 7.2.14 Observe that the argument given in (7.2.9) actually shows that if $D = \sum_{i=1}^{N} c_i$ in the cellular chain complex of $\tilde{X}$ then $N \geqslant \text{Area}_a(w)$ (where the c_i are 2-cells). Deduce from (7.2.12) that $\Phi_{G_{p,q}}^{\text{ab}}(n)$ is $\succeq n^{2\log_2 \frac{2p}{q}}$.

[29] The conjugation action of F on K induces an action of Γ on $K/[K, K]$, hence the module structure.

8 Other Decision Problems and Measures of Complexity

8.1 Alternative Analyses of the Word Problem

We saw in Section 1 that the Dehn function measures one's likelihood of success when one mounts a direct attack on the word problem for a finitely presented group. But there are other interesting ways to measure the complexity of the word problem. For example, instead of focusing on the area of van Kampen diagrams one might focus on some other aspect of their geometry, such as their diameter or the radius of the largest ball in the interior of the diagram. One might also bound the length of the intermediate words that arise during the process of applying relations to reduce a null-homotopic word to the empty word – "filling length". In Chapters 4 and 5 of [56] Gromov discusses many measures of complexity such as these, and there has been some interesting work on their interdependency (e.g. [46], [18], and [50]). Let me describe the most widely studied of these alternatives, which relates to the diameter of filling-discs in Riemannian manifolds.

Definition 8.1.1 *Let $\langle \mathcal{A} \mid \mathcal{R} \rangle$ be a finite presentation for the group Γ. Let w be a word that equals 1 in Γ and let D be a van Kampen diagram for w. Let p be the basepoint of D. Endow the 1-skeleton of D with a path metric ρ that gives each edge length 1. The diameter of w is defined by*

$$\mathrm{diam}(w) := \min_D \max_q \{\rho(p,q) \mid q \textit{ a vertex of } D\}.$$

The (unreduced) isodiametric function of $\langle \mathcal{A} \mid \mathcal{R} \rangle$ is

$$\Psi(n) := \max_{|w| \leqslant n} \mathrm{diam}(w).$$

The $\simeq$ equivalence class of Ψ depends only on Γ (see [46]) and is denoted Ψ_Γ.

Isodiametric functions turn out to be as unconstrained in nature as Dehn functions (3.1.11), see [95]. They can be interpreted in the following purely algebraic manner.

Proposition 8.1.2 $\mathrm{diam}(w) = \min_\Pi \max |x_i|$, *where the minimum is taken over all free equalities of the form*

$$w = \prod_{i=1}^{N} x_i^{-1} r x_i.$$

Exercises 8.1.3 (i) Deduce this proposition from the constructions in Section 4.

(ii) Use the diagrams constructed in (7.1.3) to show that if a group Γ admits a combing with the asynchronous fellow-traveller property, then $\Psi_\Gamma(n) \simeq n$.

(iii) Prove that $\Psi_\Gamma \preceq \delta_\Gamma$ for all finitely presented groups.

Steve Gersten and Daniel Cohen (independently) proved that for any group one can find constants $A, B > 0$ such that $\delta_\Gamma(n) \leqslant A^{B^{\Psi(n)}}$, and it is conjectured that in reality there is a single exponential bound. The relationship between Ψ_Γ and δ_Γ is complicated by the fact that in general the minima in the definitions of these functions will not be attained on the same family of diagrams: if one proceeds as in Definition 8.1.1 but quantifies only over least-area diagrams, then one obtains a function $\Psi_\Gamma^{\rm ma}$ that in general is $\prec \Psi_\Gamma$.

8.1.4 Extrinsic Solutions to the Word Problem. In general, invariants based entirely on the geometry of van Kampen diagrams cannot give a full and accurate measure of the complexity of the word problem in a group because there might exist algorithms that require *extrinsic* structure that cannot be seen in a presentation. For example, one can solve the word problem for $B_1 = \langle x_0, x_1 \mid x_1^{-1}x_0x_1 = x_1^2\rangle$ in polynomial time by looking at the orbit of $\frac{1}{3} \in \mathbb{R}$ under the action $B_1 \to \mathrm{Aff}(\mathbb{R})$ described following (3.2.1), and yet $\delta_{B_1}(n) \simeq 2^n$.

If there is an embedding $\Gamma \hookrightarrow \hat{\Gamma}$ into a group whose Dehn function is smaller than that of Γ then one can apply the solution to the word problem in $\hat{\Gamma}$ to solve the word problem in Γ. Examples of this phenomena are described in [8], [47] and [22] page 487. Remarkably, in [13] Birget, Ol'shanskii, Rips and Sapir prove that such embeddings take full account of the complexity of the word problem in a precise sense that includes the following statement: the word problem of a finitely generated group G is an NP problem if and only if G is a subgroup of a finitely presented group that has a polynomial Dehn function.

8.2 Other Decision Problems

In this article we are concentrating on the word problem, but I should say a few words about the complexity of the other basic decision problems in group theory.

We fix a group Γ with a finite generating set $\mathcal{A}$. In order to solve the word problem one must decide which words in the letters $\mathcal{A}^{\pm 1}$ equal $1 \in \Gamma$. Two natural generalisations of this problem are:

(1) *The Membership Problem (Relative Word Problem).* Instead of determining which words represent elements of the trivial subgroup, one is asked for an algorithm that decides which words represent elements of the subgroup $H \subset \Gamma$ generated by a specified finite subset of Γ.

(2) *The Conjugacy Problem.* Instead of determining which words represent elements conjugate to the identity, one is asked for an algorithm that decides which pairs of words represent conjugate elements of Γ.

Just as solving the word problem in Γ amounts to finding discs with a specified boundary loop in a closed manifold M with $\pi_1 M = \Gamma$, so the conjugacy problem amounts to finding annuli whose boundary is a specified pair of loops (minimizing the thickness of the annulus corresponds to bounding the length of the conjugating element). In the same vein, the membership problem corresponds to determining which paths can be homotoped (rel endpoints) into a given subspace of M.

There are various constructions connecting the word, conjugacy and membership problems — see [73] and [9]. The following fibre product construction provides a particularly nice example as it can be modelled readily in geometric settings.

Exercise 8.2.1 Let $\Gamma = \langle \mathcal{A} \mid \mathcal{R} \rangle$ be a finitely presented group and let $D \subset F(\mathcal{A}) \times F(\mathcal{A})$ be the subgroup $\{(w, w') \mid w = w' \text{ in } \Gamma\}$. Show that D is finitely generated. Explain why solving the word problem for Γ is equivalent to solving the membership problem for D. Show that if one cannot solve the word problem in Γ then one cannot solve the conjugacy problem in D. (Hint: Fix $r \in \mathcal{R}$. Given a word w in the generators of $F(\mathcal{A}) \times \{1\}$, express the element $w^{-1}(r, r)w$ as a word in your chosen generators of D. When is the word you have created conjugate to (r, r) in D?)

If Γ is infinite then the group D in the above exercise is not finitely presentable (see [57]). For finitely presented examples and variations of a more geometric nature, see [9].

Remark 8.2.2 The conjugacy problem is considerably more delicate than the word problem in general. For example, in contrast to the fact that the complexity of the word problem for a group remains essentially unchanged when one passes to a subgroup or overgroup of finite index (1.3.5), Collins and Miller [32] constructed pairs of finitely presented groups $H \subset G$ such that $|G/H| = 2$ but H has a solvable conjugacy problem while G does not. They also show that one can arrange for G to have a solvable conjugacy problem when H does not.

The Isomorphism Problem. Roughly speaking, the isomorphism problem asks for an algorithm that will decide which finite presentations drawn from a specified list define isomorphic groups. The difficulty of this problem depends very much on the nature of the groups being presented. For example, Zlil Sela [96] proved that if one is given the knowledge that all of the groups being presented are the fundamental groups of closed negatively curved manifolds, then there is an algorithm that one can run to decide which of the groups are isomorphic. In contrast, it is unknown if there exists such an algorithm when one weakens the curvature condition to allow non-positively curved manifolds. Indeed there are very few natural contexts in which the isomorphism problem has been solved. (Note that in order to solve the isomorphism problem in a given class of groups it is not enough to have an algorithm that determines which presentations give the trivial group; for example, there is an algorithm to decide whether presentations of automatic groups determine the trivial group (chapter 5 of [42]) but this does not lead to a solution of the isomorphism problem in this class of groups.)

The following construction illustrates how HNN extensions can be used to translate word problems into other sorts of decision problems.

Exercise 8.2.3 Let $\Gamma = \langle \mathcal{A} \mid \mathcal{R} \rangle$ be a finitely presented group that is not free. Suppose that $\mathcal{A} = \{a_1, \dots, a_n\}$ where each a_i has infinite order in Γ (this can be arranged by replacing Γ with $\Gamma * \mathbb{Z}$ if necessary). Consider the following sequence of finite presentations indexed by words $w \in F(\mathcal{A})$:

$$G_w = \langle a_1, t_1, \dots, a_n, t_n \mid \mathcal{R}, t_i^{-1} a_i t_i = w \text{ for } i = 1, \dots, n \rangle.$$

Show that Γ_w is a free group if and only if $w = 1$ in Γ.

Assuming that there exists a group with an unsolvable word problem, use this construction (or a variation on it) to show that there exist (recursive) classes of finite presentations such that there are no algorithms to decide which of the groups presented are free, are torsion-free, contain $\mathbb{Z}^2$ (or any other specified subgroup), can be generated by 3 elements, or admit a faithful representation into $\mathrm{SL}(n, \mathbb{Z})$.

8.3 Subgroup Distortion

Following Gromov [56], we define the distortion of a pair of finitely generated groups $H \subset \Gamma$ to be the function $\rho : \mathbb{N} \to \mathbb{N}$, where $\rho(n)$ is the radius of the set of vertices in the Cayley graph of H that are a distance at most n from the identity in Γ. (One shows that, up to $\simeq$ equivalence, this function does not depend on the choice of generating sets.)

If Γ has a solvable word problem, then the membership problem for $H \subset \Gamma$ is solvable if and only if the *distortion* of H in Γ is a recursive function.

Examples 8.3.1 (i) If $\phi \in \mathrm{GL}(r, \mathbb{Z})$ has an eigenvalue of absolute value greater than 1, then $\mathbb{Z}^r$ is exponentially distorted in $\mathbb{Z}^r \rtimes_\phi \mathbb{Z}$.

(ii) Let G_c be as in (3.1.7). In [19] I proved that for all positive integers $a > b$ the distortion of G_b in $G_a *_{\langle z \rangle} G_b$, the group formed by amalgamating G_a and G_b along their centres, is $\simeq n^{\frac{a}{b}}$. In [85] Osin proves that one can also obtain arbitrary positive rational exponents a/b by considering subgroups of finitely generated nilpotent groups.

(iii) Let $G_{p,q}$ be as in (3.1.10). In [15] Brady and I proved that the distortion function of the torus subgroup $\langle a, b \rangle$ in $G_{p,q}$ is equivalent to n^α, where $\alpha = \log_2(2p/q)$.

Ol'shanskii and Sapir have established comprehensive results, analogous to Theorem 3.1.11, concerning the possible distortion functions of finitely presented subgroups — see [82], [84].

See [22] page 507 for an interpretation of subgroup distortion in terms of Riemannian geometry as well as a connection between subgroup distortion and Dehn functions. See [43] for a discussion of relative Dehn functions.

A Geometric Realisations of Finitely Presented Groups

This appendix contains a brief description of some of the basic constructions of geometric group theory. There are two main (inter-related) strands in geometric group theory: one seeks to understand groups by studying their actions on appropriate spaces, and one seeks understanding from the intrinsic geometry of (discrete, finitely generated) groups endowed with word metrics. We begin by introducing the latter approach.

A.1 Finitely Generated Groups and Quasi-Isometries

The following constructions allow one to regard finitely generated groups as geometric objects.

A.1.1 Word Metrics and Cayley Graphs Given a group Γ with generating set $\mathcal{A}$, the first step towards realizing the intrinsic geometry of the group is to give Γ the *word metric* associated to $\mathcal{A}$: this is the metric obtained by defining $d_{\mathcal{A}}(\gamma_1, \gamma_2)$ to be the shortest word in the letters $\mathcal{A}^{\pm 1}$ that equals $\gamma_1^{-1}\gamma_2$ in Γ. The action of Γ on itself by left multiplication gives an embedding $\Gamma \to \mathrm{Isom}(\Gamma, d_{\mathcal{A}})$. (The action of $\gamma_0 \in G$ by right multiplication $\gamma \mapsto \gamma\gamma_0$ is an isometry only if γ_0 lies in the centre of Γ.)

The *Cayley graph*[30] of Γ with respect to $\mathcal{A}$, denoted $\mathcal{C}_{\mathcal{A}}(\Gamma)$, has vertex set Γ and has an edge connecting γ to γa for every $\gamma \in \Gamma$ and $a \in \mathcal{A}$. The edges of $\mathcal{C}_{\mathcal{A}}(\Gamma)$ are endowed with local metrics in which they have unit length, and $\mathcal{C}_{\mathcal{A}}(\Gamma)$ is turned into a geodesic space by defining the distance between each pair of points to be equal to the length of the shortest path joining them.

The word metrics associated to different finite generating sets $\mathcal{A}$ and $\mathcal{A}'$ of Γ are Lipschitz equivalent, i.e. there exists $\ell \geqslant 1$ such that $\frac{1}{\ell} d_{\mathcal{A}}(\gamma_1, \gamma_2) \leqslant d_{\mathcal{A}'}(\gamma_1, \gamma_2) \leqslant \ell\, d_{\mathcal{A}}(\gamma_1, \gamma_2)$ for all $\gamma_1, \gamma_2 \in \Gamma$. One sees this by expressing the elements of $\mathcal{A}$ as words in the generators $\mathcal{A}'$ and vice versa – the constant ℓ is the length of the longest word in the dictionary of translation.

The Cayley graphs associated to different finite generating sets are not homeomorphic in general, but they are quasi-isometric in the following sense.

Definition A.1.2 *A (not necessarily continuous) map $f : X \to X'$ between metric spaces is called a* quasi-isometry *if there exist constants $\lambda \geqslant 1, \epsilon \geqslant 0, C \geqslant 0$ such that every point of X' lies in the C-neighbourhood of $f(X)$ and*

$$\frac{1}{\lambda} d(x, y) - \epsilon \leqslant d(f(x), f(y)) \leqslant \lambda d(x, y) + \epsilon$$

for all $x, y \in X$.

---

[30] This graph was introduced by Arthur Cayley in 1878 to study "the quasi-geometrical" nature of (in his case, finite) groups. It played an important role in the seminal work of Max Dehn (1910) who gave it the name Gruppenbild.

Exercises A.1.3 (i) If there exists a quasi-isometry $X \to X'$ then X and X' are said to be quasi-isometric. Prove that being quasi-isometric is an equivalence relation on any set of metric spaces.

(ii) Show that if $\mathcal{A}$ and $\mathcal{A}'$ are finite generating sets for Γ, then $(\Gamma, d_{\mathcal{A}})$, $\mathcal{C}_{\mathcal{A}}(\Gamma)$ and $\mathcal{C}_{\mathcal{A}'}(\Gamma)$ are quasi-isometric.

(iii) When is a homomorphism between finitely generated groups a quasi-isometry?

Since the quasi-isometry type of a finitely generated group does not depend on a specific choice of generators, statements such as "the finitely generated group Γ is quasi-isometric to the metric space Y" or "the finitely generated groups Γ_1 and Γ_2 are quasi-isometric" are unambiguous.

One may view the inclusion $\Gamma \hookrightarrow \mathcal{C}_{\mathcal{A}}(\Gamma)$ in the following light: Γ acts by isometries on $\mathcal{C}_{\mathcal{A}}(\Gamma)$, the action of $\gamma_0 \in \Gamma$ sending the edge with label $a \in \mathcal{A}$ emanating from the vertex γ to the edge labelled a emanating from the vertex $\gamma_0\gamma$, and $\Gamma \hookrightarrow \mathcal{C}_{\mathcal{A}}(\Gamma)$ is the map $\gamma \mapsto \gamma \cdot 1$. This is a simple instance of the important observation that quasi-isometries arise naturally from group actions (see [22] page 140).

Proposition A.1.4 (The Švarc-Milnor Lemma) *If a group Γ acts properly and cocompactly by isometries on a length space X, then for every choice of basepoint $x_0 \in X$ the map $\gamma \mapsto \gamma.x_0$ is a quasi-isometry.*

The fundamental group of any (locally simply-connected) space acts by deck transformations on the universal covering. If the space is a compact geodesic space and the universal covering is endowed with the induced length metric ([22] page 42), then this action is proper, cocompact and by isometries. Thus we have:

Corollary A.1.5 *The fundamental group of any closed Riemannian manifold M is quasi-isometric to the universal covering $\tilde{M}$.*

We note one other corollary of the Švarc-Milnor Lemma:

Corollary A.1.6 *If X_1 and X_2 are length spaces and there is a finitely-generated group Γ that acts properly and cocompactly by isometries on both X_1 and X_2, then X_1 and X_2 are quasi-isometric.*

Dehn functions behave well with respect to quasi-isometries (see [4] and compare with Proposition 1.3.3 above and pages 143 and 415 of [22]).

Proposition A.1.7 *If Γ is a finitely presented group and Γ' is a finitely generated group quasi-isometric to Γ, then Γ' is also finitely presented and the Dehn functions of Γ and Γ' are $\simeq$ equivalent.*

By combining this proposition with the preceding corollaries and the Filling Theorem we obtain:

Theorem A.1.8 *If the universal coverings of two closed, smooth, Riemannian manifolds M_1 and M_2 are quasi-isometric, then the isoperimetric functions $\mathrm{Fill}_0^{M_1}$ and $\mathrm{Fill}_0^{M_2}$ are $\simeq$ equivalent.*

One can prove this result more directly by following Alonso's proof of (A.1.7) using the combinatorial approximation techniques developed in Section 5.

A.2 Realising the Geometry of Finite Presentations

We now focus on finitely presented groups. The following category of complexes and maps is more rigid than the CW category and lends itself well to arguments such as those that we saw in the section on Van Kampen's Lemma. The discussion here follows that of Appendix I.8.A in [22].

A.2.1 Combinatorial Complexes These complexes are topological objects with a specified combinatorial structure. They are defined by a recursion on dimension; the definition of an open cell is defined by a simultaneous recursion. If K_1 and K_2 are combinatorial complexes, then a continuous map $K_1 \to K_2$ is said to be *combinatorial* if its restriction to each open cell of K_1 is a homeomorphism onto an open cell of K_2.

A combinatorial complex of dimension 0 is simply a set with the discrete topology; each point is an open cell. Having defined $(n-1)$-dimensional combinatorial complexes and their open cells, one constructs n-dimensional combinatorial complexes as follows.

Take the disjoint union of an $(n-1)$-dimensional combinatorial complex $K^{(n-1)}$ and a family $(e_\lambda \mid \lambda \in \Lambda)$ of copies of closed n-dimensional discs. Suppose that for each $\lambda \in \Lambda$ a homeomorphism is given from ∂e_λ (a sphere) to an $(n-1)$-dimensional combinatorial complex S_λ, and that a combinatorial map $S_\lambda \to K^{(n-1)}$ is also given; let $\phi_\lambda : \partial e_\lambda \to K^{(n-1)}$ be the composition of these maps. Define K to be the quotient of $K^{(n-1)} \cup \coprod_\Lambda e_\lambda$ by the equivalence relation generated by $t \sim \phi_\lambda(t)$ for all $\lambda \in \Lambda$ and all $t \in \partial e_\lambda$. Then K, with the quotient topology, is an n-dimensional combinatorial complex whose open cells are the (images of) open cells in $K^{(n-1)}$ and the interiors of the e_λ.

In the case $n = 2$, if the circle S_λ has k 1-cells then e_λ is called a k-gon.

A.2.2 The Standard 2-Complex $K(\mathcal{A} : \mathcal{R})$ Associated to any group presentation $\langle \mathcal{A} \mid \mathcal{R} \rangle$ one has a 2-complex $K = K(\mathcal{A} : \mathcal{R})$ that is compact if and only if the presentation is finite. K has one vertex and it has one edge ε_a (oriented and labelled a) for each generator $a \in \mathcal{A}$; thus edge loops in the 1-skeleton of K are in 1–1 correspondence with words in the alphabet $\mathcal{A}^{\pm 1}$: the letter a^{-1} corresponds to traversing the edge ε_a in the direction opposite to its orientation, and the word $w = a_1 \dots a_n$ corresponds to the loop that is the concatenation of the directed edges $a_1, \dots, a_n$; one says that w *labels* this loop. The 2-cells e_r of K are indexed by the relations $r \in \mathcal{R}$; if $r = a_1 \dots a_n$ (as a reduced word) then e_r is attached along the loop labelled $a_1 \dots a_n$. The map that sends the homotopy class of ε_a to $a \in \Gamma$ gives an isomorphism $\pi_1 K(\mathcal{A} : \mathcal{R}) \cong \Gamma$ (by the Seifert-van Kampen theorem).

Γ acts on the universal covering $\widetilde{K}$ of $K(\mathcal{A}:\mathcal{R})$ by deck transformations and there is a natural Γ-equivariant identification of the Cayley graph $\mathcal{C}_{\mathcal{A}}(\Gamma)$ with the 1-skeleton of $\widetilde{K}$: fix a base vertex $v_0 \in \widetilde{K}(\mathcal{A}:\mathcal{R})$, identify $\gamma.v_0$ with γ, and identify the edge of $\mathcal{C}_{\mathcal{A}}(\Gamma)$ labelled a issuing from γ with the (directed) edge at $\gamma.v_0$ in the pre-image of ε_a. This identification is label-preserving: for all words w and all $\gamma \in \Gamma$, there is a unique edge-path labelled w beginning at $\gamma \in \mathcal{C}_{\mathcal{A}}(\Gamma)$ and the image of this path in $\tilde{K}$ is the lift at $\gamma.v_0$ of the loop in $K(\mathcal{A}:\mathcal{R})$ labelled w.

Exercise A.2.3 Prove that if $\mathcal{A}$ is finite and w is a reduced word in which a and a^{-1} both occur exactly once, for every $a \in \mathcal{A}$, then $K = K(\mathcal{A}:w)$ is obtained from a closed surface by gluing together a finite set of points.

A.3 4-Manifolds Associated to Finite Presentations

Proposition A.3.1 *Every finitely presented group is the fundamental group of a closed 4-dimensional manifold.*

We indicate two proofs of this proposition, leaving the details to the reader.

Exercises A.3.2 (i) Given a presentation $\langle a_1, \dots, a_n \mid r_1, \dots, r_m \rangle$, consider the compact 4-manifold obtained by taking the connected sum W of n copies of $\mathbb{S}^1 \times \mathbb{S}^3$ and identify $\pi_1 W$ with the free group on $\{a_1, \dots, a_n\}$. Remove open tubular neighbourhoods about m disjoint embedded loops in W whose homotopy classes correspond to the relators $r_i \in \pi_1 W$. Let W' be the resulting manifold with boundary. Use the Seifert-van Kampen theorem to show that by attaching m copies of $\mathbb{S}^2 \times \mathbb{D}^2$ to W' along $\partial W'$ one obtains a closed manifold whose fundamental group is $\langle a_1, \dots, a_n \mid r_1, \dots, r_m \rangle$.

(ii) Show that if $n \geqslant 4$ then one can embed any compact combinatorial 2-complex in $\mathbb{R}^4$ by a piecewise linear map. Apply this construction to $K(\mathcal{A}:\mathcal{R})$ and consider the boundary M of a regular neighbourhood. Argue that the natural map $\pi_1 M \to \langle \mathcal{A} \mid \mathcal{R} \rangle$ is an isomorphism if $n \geqslant 5$.

By performing constructions of the above type more carefully one can force the manifold to have additional structure. For example, in [52] Bob Gompf proves:

Theorem A.3.3 *Every finitely presented group is the fundamental group of a closed symplectic 4-manifold.*

A.4 Obtaining Presentations from Group Actions

Whenever one realises a group as the fundamental group of a (semi-locally simply-connected) space one has the action of the group by deck transformations on the universal covering of the space. Thus the constructions of $K(\mathcal{A}:\mathcal{R})$ and the manifolds considered above may be viewed as means of constructing group actions out of presentations. The following theorem shows that, conversely, group actions give rise to presentations.

Theorem A.4.1 *Let X be a topological space, let Γ be a group acting on X by homeomorphisms, and let $U \subset X$ be an open subset such that $X = \Gamma.U$.*

(1) *If X is connected, then the set $S = \{\gamma \in \Gamma \mid \gamma.U \cap U \neq \emptyset\}$ generates Γ.*

(2) *Let $\mathcal{A}_S$ be a set of symbols a_s indexed by S. If X and U are both path-connected and X is simply connected, then $\Gamma = \langle \mathcal{A}_S \mid \mathcal{R} \rangle$, where*

$$\mathcal{R} = \{a_{s_1} a_{s_2} a_{s_3}^{-1} \mid s_i \in S;\ U \cap s_1.U \cap s_3.U \neq \emptyset;\ s_1 s_2 = s_3 \text{ in } \Gamma\}.$$

Corollary A.4.2 *If a group Γ acts by isometries on a complete Riemannian manifold M, and if every point of M is a distance less than r from a certain orbit $\Gamma \cdot p$, then Γ can be presented as $\Gamma = \langle \mathcal{A} \mid \mathcal{R} \rangle$ where $\mathcal{A}$ is the set of elements $a \in \Gamma$ such that $d(p, \gamma \cdot p) < 2r$ and $\mathcal{R}$ is the set of words in the letters $\mathcal{A}^{\pm 1}$ that have length at most 3 and are equal to the identity in Γ.*

Proof Apply the theorem with U the open ball of radius r about p. □

The above theorem has a long history. In this form it is due to Murray Macbeath [68]. See [22] page 136 for a proof and further information.

Exercises A.4.3 Establish the following geometric characterisation of finitely presented groups: a group is finitely presented if and only if it acts properly and cocompactly by isometries on a simply-connected geodesic space.

Give an example to show that part (2) of the above theorem can fail if X is not simply connected.

B Length Spaces

For the benefit of the reader unfamiliar with non-Riemannian length spaces we list some of the basic vocabulary of the subject.

Length Metrics.

Definition B.0.1 *Let X be a metric space. The* length *$l(c)$ of a curve $c : [a, b] \to X$ is*

$$l(c) = \sup_{a = t_0 \leqslant t_1 \leqslant \ldots \leqslant t_n = b} \sum_{i=0}^{n-1} d(c(t_i), c(t_{i+1})),$$

where the supremum is taken over all possible partitions (no bound on n) with $a = t_0 \leqslant t_1 \leqslant \ldots \leqslant t_n = b$.

$l(c)$ is either a non-negative number or it is infinite. The curve c is said to be *rectifiable* if its length is finite, and it is called a *geodesic*[31] if its length is equal to the distance between its endpoints.

A *triangle* Δ in a metric space consists of three points x, y, z (the vertices) and a choice of geodesic connecting each pair of these points.

A (connected) *length space* is a metric space X in which every pair of points $x, y \in X$ can be joined by a rectifiable curve and $d(x, y)$ is equal to the infimum of the length of rectifiable curves joining them; X is called a *geodesic space* if this infimum is always attained, i.e. each pair of points $x, y \in X$ can be joined by a geodesic. A general form of the Hopf-Rinow Theorem (see [6] or [22]) states that if a length space is complete, connected and locally compact, then it is a geodesic space (and all closed balls in it are compact).

Upper Curvature Bounds. Let M_k^2 denote the complete simply-connected 2-manifold of constant sectional curvature $k \in \mathbb{R}$. (If $k = 0$ then M_k^2 is the Euclidean plane; if $k < 0$ then M_k^2 is the hyperbolic plane with the metric scaled by a factor of $1/\sqrt{-k}$; and if $k > 0$ then M_k^2 is $\mathbb{S}^2$ with the metric scaled by $1/\sqrt{k}$.

A *triangle* Δ in a metric space consists of three points x_1, x_2, x_3 (the vertices) and a choice of geodesic connecting each pair of these points.

A geodesic space X is said to have *curvature* $\leqslant k$ if every point $x \in X$ has a neighbourhood in which all triangles Δ satisfy the following property: the distance from each vertex of Δ to the midpoint of the opposite side is no greater than the corresponding distance in a triangle $\overline{\Delta} \subset M_k^2$ that has the same edge lengths as Δ. This definition is due to A.D. Alexandrov.

We refer the reader to [22] for a comprehensive introduction to (singular) spaces with upper curvature bounds.

Pull-Back Length Metrics. Let D be a topological space. Associated to any continuous map $f : D \to X$ to a metric space one has the length pseudo-metric on D: the length of each curve in D is defined to be the length of its image under f, and the distance between two points of D is defined to be the infimum of the lengths of paths connecting them. We write (D, d_f) to denote the length space obtained by forming the quotient of this pseudo-metric space by the relation that identifies points that are a distance 0 apart. In general one can say little about the underlying space of (D, d_f); it certainly need not be homeomorphic to D.

If X is a smooth Riemannian manifold and $f : D \to X$ is a least-area disc with piecewise geodesic boundary, then (D, d_f) will be a singular disc and its curvature will be bounded above by the sectional curvature of X; if $f|_{\partial D}$ is injective, then (D, d_f) will actually be a disc. It can also be that (D, d_f) is a disc when f is not injective, for example if f is the map $z \mapsto z^2$ from the unit

---

[31] This differs from the standard usage in differential geometry, where being geodesic is a local concept. For this reason, some authors use the term "length-minimizing geodesic" in the context of length spaces.

disc to the complex plane, then (D, d_f) is the metric completion of the connected 2-fold covering of the punctured unit disc.

C A Proof of the Cellulation Lemma

This appendix contains a proof of the following technical result that was needed in Section 5. Recall that a *singular disc* is a space homeomorphic to the underlying space of a singular disc diagram, as defined in (4.1).

C.0.1 Cellulation Lemma. *Let D be a length space homeomorphic to a (perhaps singular) 2-disc, and suppose that D is ε-filled by a set Σ of cardinality N. Then there exists a combinatorial 2-complex Φ, homeomorphic to the standard 2-disc, and a continuous map $\phi : \Phi \to D$ such that:*

(1) *Φ has less than $8N$ faces (2-cells) and each is a k-gon with $k \leqslant 12$;*

(2) *the restriction of ϕ to each 1-cell in Φ is a path of length at most 2ε;*

(3) *$\phi|_{\partial\Phi}$ is a monotone parameterisation of ∂D and $\Sigma \cap \partial D$ lies in the image of the 0-skeleton of $\partial\Phi$.*

For convenience we rescale the metric on D and assume that $\varepsilon = 1$. To avoid complicating the terminology, we also assume that D is a non-singular disc (the concerned reader will have little difficulty in making the adjustments needed in the general case). We fix a set Σ of cardinality N that 1-fills D and define $\Sigma_0 = \Sigma \cap \partial D$ and $\Sigma_1 = \Sigma \smallsetminus \Sigma_0$.

C.1 Reducing to the Case of Thin Discs

Our aim in the first stage of the proof is to reduce to the case where $\Sigma = \Sigma_0$. We shall do this by cutting D open along a certain graph whose vertex set has cardinality less than $2N$ and includes Σ. To this end, we view ∂D as a graph $\mathcal{G}_0$ with vertex set Σ_0 and 1-cells the closures of the connected components of $\partial D \smallsetminus \Sigma_0$.

Since every point of the connected space D lies in the 1-neighbourhood of Σ, the open neighbourhoods of radius 1 about Σ_0 and Σ_1 cannot be disjoint. Hence there exists $s \in \Sigma_1$ and $s' \in \Sigma_0$ with $d(s, s') < 2$. Choose a geodesic $[s, s']$ and consider a minimal subarc $[s, v]$ with $v \in \mathcal{G}_0$. We augment $\mathcal{G}_0$ (which is ∂D subdivided) by adding s and v as vertices and adding $[s, v]$ as a new edge (if v is not a vertex of $\mathcal{G}_0$ then its introduction will also subdivide one of the existing edges). Call the new graph $\mathcal{G}_0'$ and define $\Sigma_0' = \Sigma_0 \cup \{s\}$.

By repeating the above argument with Σ_0' in place of Σ_0, and $\mathcal{G}_0'$ in place of $\mathcal{G}_0$, we obtain a connected graph with at most $|\Sigma_0| + 4$ vertices including Σ_0 and two elements of Σ_1. We iterate this argument a further $|\Sigma_1| - 2$ times to obtain a connected graph $\mathcal{G}$ whose vertex set consists of Σ and at most $2|\Sigma_1|$ other vertices; the important point is that this graph has less than $2N$ vertices in total, and less than $2N$ edges. Note that the edges of $\mathcal{G}$ all have length at most 2, that $E := D \smallsetminus \mathcal{G}$ is homeomorphic to an open 2-disc, and that $T := \mathcal{G} \smallsetminus \partial D$ is a forest (i.e. it is simply-connected, but not necessarily connected).

We now focus our attention on E, which we endow with the induced path metric from D. Let Δ be the space obtained by completing this metric. Δ is homeomorphic to a 2-disc; intuitively speaking, it is obtained by cutting D open along the branches of T (cutting along each edge of T forms two edges in the boundary of Δ). The inclusion $E \hookrightarrow D$ extends continuously to a map $\pi : \Delta \to D$ that preserves the lengths of all curves and sends (a monotone parameterization of) $\partial\Delta$ onto the boundary cycle of E in $\mathcal{G}$; we endow $\partial\Delta$ with the combinatorial structure induced from this identification. Thus Δ is a topological 2-disc endowed with a length metric such that $\partial\Delta$ is the concatenation of less than $4N$ geodesic segments, each of length at most 2. Moreover, every point of Δ is a distance at most 1 from $\partial\Delta$. This completes the first stage of the proof.

Definition C.1.1 *A* singular disc of weight n *consists of a singular disc* Δ *and* n *distinguished points (*vertices*)* $x_1 = f(t_1), \ldots, x_n = f(t_n)$ *in cyclic order on the boundary cycle* $f : \mathbb{S}^1 \to \partial\Delta$*; the restriction of* f *to the arc joining* t_i *to* t_{i+1} *(indices mod* n*) is required to be a geodesic of length at most* 2*; the images of these arcs are called* facets*.* Δ *is said to be* thin *if every point is a distance less than 1 from* $\partial\Delta$*.*

A partition *of* Δ *is a continuous map* $\phi : \Phi \to \Delta$*, where* Φ *is a combinatorial 2-complex that is homeomorphic to the standard disc and* $\phi|_{\partial\Phi}$ *is a monotone parameterisation of* f *sending vertices to vertices and edges to facets.*

Φ *is called a* k-partition *if each of its 2-cells is an* m*-gon with* $m \leqslant k$*. And* Φ *is said to be* admissible *if the restriction of* ϕ *to each 1-cell in* Φ *is a path of length at most* 2*. The* area *of* Φ *is the number of 2-cells in* Φ*.*

The final stage in the proof of the Cellulation Lemma is:

Proposition C.1.2 *If* $k \geqslant 12$*, then every thin singular disc of weight* n *admits a* k-partition *of area at most* $2n - 8$*.*

Before turning to the proof of this proposition, let us see how it implies the Cellulation Lemma.

End of the proof of the Cellulation Lemma. In the first stage of the proof we showed that if a disc can be ε-filled with a set of cardinality N then one can construct in D a graph $\mathcal{G}$ with at most $2N$ vertices so that the edges of the graph have length less than 2ε and the space obtained by cutting D open along the forest $T = \mathcal{G} \smallsetminus \partial D$ is a thin disc X of weight less than $4N$. The natural map $\pi : \Delta \to D$ is length-preserving.

The above proposition furnishes a 12-partition $\phi_0 : \Phi_0 \to \Delta$ of area at most $8N - 8$. Define Φ to be the combinatorial complex obtained by taking the quotient of Φ' by the equivalence relation that identifies the pair of edges in the pre-image of each edge of T in the obvious manner. Φ is a disc whose area (number of 2-cells) is the same that of Φ_0. The map $\phi : \Phi \to D$ induced by $\pi \circ \phi_0 : \Phi_0 \to D$ satisfies the requirements of the Cellulation Lemma. $\square$

C.2 Surgery on Thin Discs

We shall prove Proposition C.1.2 by induction on n, the weight of the singular disc being filled. In this induction we shall need the following surgery operation.

Let Δ be a singular disc of weight n with boundary cycle $f : \mathbb{S}^1 \to \partial\Delta$. Given two vertices $x, y \in \partial\Delta$ one can cut Δ along a geodesic $[x, y]$ to form two new singular discs. To do this, first note that one can choose $[x, y]$ so that its intersection with each facet of $\partial\Delta$ is a single arc, because given the first and last points of intersection of an arbitrary geodesic $[x, y]'$ with a facet, one can replace the corresponding subarc of $[x, y]'$ with a subarc of the facet. Having chosen $[x, y]$ in this way, express y as $f(t)$ and proceed in the positive direction around $\mathbb{S}^1$ from t to the first value t' such that $f(t') = x$; let α denote this arc from t to t' and call the complementary arc β'.

The first of the two singular sub-discs into which we cut Δ is that whose boundary cycle is the concatenation of $f|_\alpha$ and $[x, y]$. The boundary cycle of the second sub-disc is the concatenation of $f|_\beta$ and $[y, x]$. We subdivide $[x, y]$ into the minimal possible number of sub-arcs of length less than 2 and define these sub-arcs to be facets of our two new singular discs.

The reader should have no difficulty in verifying:

Lemma C.2.1 *In the notation of the preceding paragraph: if Δ is thin then the singular discs obtained by surgery are thin; and if $d(x, y) < 4$, then the sum of the weights of the new singular discs is at most $n + 4$.*

In the course of the proof of Proposition C.1.2 we shall require the following fact.

Exercise C.2.2 Let $X = U_1 \cup U_2 \cup U_3 \cup U_4$ be a metric space. Assume that each of the sets U_i is path-connected, that $d(U_i, U_j) > 0$ when $|i - j| = 2$, and that $U_i \cap U_j \neq \emptyset$ otherwise. Construct a surjective homomorphism $\pi_1 X \to \mathbb{Z}$. (Hint: Consider the map to $\mathbb{R}/\mathbb{Z}$ that is constant on $X \smallsetminus U_2$ and is given on U_2 by $x \mapsto d(x, U_1)/(d(x, U_1) + d(x, U_3))$.)

The Proof of Proposition C.1.2. Let Δ be a singular disc of weight n that is thin. We proceed by induction on n. If $n \leqslant k$ there is nothing to prove.

Assuming $n \geqslant 12$, we express the boundary cycle $f : \mathbb{S}^1 \to \Delta$ as the concatenation of four subpaths, namely the first three facets taken together, the next three facets, then the next three, and then the remaining $n - 9$ facets. Define U_1, U_2, U_3, U_4 to be the closed neighbourhoods of radius 1 about the images of these four arcs. The union of these neighbourhoods is the whole of Δ (because it is assumed to be thin). The U_i cannot satisfy the hypotheses of the preceding exercise because Δ is simply connected. Therefore $U_i \cap U_j \neq \emptyset$ for some $i - j = 2$. (Here we need the fact that the metric on Δ is a path metric in order to know that the U_i are path-connected.)

Since U_i and U_j intersect, one of the vertices along our i-th arc, say x, is a distance at most 4 from one of the vertices along our j-th arc, say y. We separate Δ by surgery along $[x, y]$. Because all four of our sub-arcs contained at least 3 facets, and because we need only divide $[x, y]$ into two facets, the weights n' and n'' of the new singular discs Δ' and Δ'' obtained by surgery are both strictly less than n. Also (see the lemma) $n' + n'' \leqslant n + 4$.

By induction, there exist admissible k-partitions $\Phi' \to \Delta'$ and $\Phi'' \to \Delta''$ whose areas are at most $2n' - 8$ and $2n'' - 8$ respectively. Let Φ be the combinatorial disc obtained by gluing Φ' and Φ'' along the pre-images of $[x, y]$ in the obvious manner. The given maps $\Phi' \to \Delta'$ and $\Phi'' \to \Delta''$ define an admissible k-partition $\Phi \to \Delta$ whose area is the sum of the areas of Φ' and Φ''. In particular the area of Φ is at most $2(n' + n'') - 16 \leqslant 2(n + 4) - 16 = 2n - 8$, so the induction is complete. $\square$

The bound $k \geqslant 12$ in Proposition C.1.2 can be improved at the expense of complicating the proof.

Acknowledgement. I first drafted a proof of the Filling Theorem (Section 4) in 1994 with the intention of saving it for a panorama such as this. I was on the faculty at Princeton then and Fred Almgren was my neighbour in Fine Hall. He answered my questions on geometric measure theory with patience and took a stimulating interest in the connections with group theory. Although he is no longer with us, I wish to record my gratitude to Fred for these conversations.

References

1. A.D. Alexandrov: A theorem on triangles in a metric space and some of its applications, *Trudy Mat. Inst. Steklov* **3**8 (1951), 5–23.

2. D.J. Allcock: An isoperimetric inequality for the Heisenberg groups, Geom. Funct. Anal. **8** (1998), 219–233.

3. F.J. Almgren Jr: 'Plateau's problem: An invitation to varifold geometry', W. A. Benjamin, Inc., New York-Amsterdam, 1966.

4. J. M. Alonso: Inégalités isopérimétriques et quasi-isométries, *C.R.A.S. Paris Série 1* **3**11 (1990), 761–764.

5. J. Alonso and M.R. Bridson: Semihyperbolic groups, *Proc. London Math. Soc.* **7**0 (1995), 56–114.

6. W. Ballmann: 'Lectures on spaces of nonpositive curvature', DMV Seminar 25, Birkhäuser Verlag, Basel, 1995.

7. G. Baumslag, S. Gersten, M. Shapiro and H. Short: Automatic groups and amalgams, *J. Pure and Appl. Alg.* **7**6 (1991), 229-316.

8. G. Baumslag, M.R. Bridson, C.F. Miller III, H. Short: Finitely presented subgroups of automatic groups and their isoperimetric functions, *J. London Math. Soc.* (2) **5**6 (1997), 292–304.

9. G. Baumslag, M.R. Bridson, C.F. Miller III, H. Short: Fibre products, nonpositive curvature and decision problems, *Comment. Math. Helv.*, **7**5 (2000), 457–477.

10. G. Baumslag, C.F. Miller III and H. Short: Isoperimetric inequalities and the homology of groups, *Invent. Math.*, **1**13 (1993), 531–560.

11. V. N. Berestovskii and I. G. Nikolaev: Multidimensional generalized Riemannian spaces, in *Geometry IV*, Encyclopedia of Mathematical Sciences, (Yu. G. Reshetnyak, ed), Springer Verlag **7**0 (1993), 165–243.

12. M. Bestvina and M. Feighn: A combination theorem for negatively curved groups, *J. Diff. Geom.* **35** (1992), 85–101.

13. J-C. Birget, A.Yu. Ol'shanskii, E. Rips, M. Sapir: Isoperimetric functions of groups and computational complexity of the word problem, preprint, http://atlas.math.vanderbilt.edu/m̃sapir/ftp/pub/bors.

14. B. H. Bowditch: A short proof that a sub-quadratic isoperimetric inequality implies a linear one, *Mich. J. Math.***4**2 (1995), 103–107.

15. N. Brady and M.R. Bridson: There is only one gap in the isoperimetric spectrum, GAFA (Geom. funct. anal.) **1**0 (2000), 1053-1070.

16. M.R. Bridson: On the geometry of normal forms in discrete groups, *Proc. London Math. Soc.* (3) **67** (1993), 596–616.

17. M.R. Bridson: Combings of semidirect products and 3-manifold groups, *Geom. Funct. Anal.* **3** (1993), 263–278.

18. M.R. Bridson: Asymptotic cones and polynomial isoperimetric inequalities, Topology **3**8 (1999), 543–554.

19. M.R. Bridson: Fractional isoperimetric inequalities and subgroup distortion, *J. Amer. Math. Soc.* **1**2 (1999), 1103–1118.

20. M.R. Bridson: Polynomial Dehn functions and the length of asynchronous automatic structures, *Proc. London Math. Soc.*, to appear.

21. M.R. Bridson and S.M. Gersten: The optimal isoperimetric inequality for torus bundles over the circle, *Quart. J. Math. Oxford* Ser. (2) **4**7 (1996), 1–23.

22. M.R. Bridson and A. Haefliger: 'Metric spaces of non-positive curvature', Grundlehren der math Wiss., 319, Springer-Verlag, Berlin–Heidelberg–New York, 1999.

23. M.R. Bridson and Ch. Pittet: Isoperimetric inequalities for the fundamental groups of torus bundles over the circle, *Geom. Dedicata* **4**9 (1994), 203–219.

24. M.R. Bridson and K. Vogtmann: On the geometry of the automorphism group of a free group, *Bull. London Math. Soc.* **2**7 (1995), 544–552.

25. K.S. Brown: 'Cohomology of groups', GTM 87, Springer-Verlag, New York (1982).

26. J. Burillo and J. Taback: The equivalence of geometric and combinatorial Dehn functions, Preprint, Tufts University, 1999.

27. J. W. Cannon: The combinatorial structure of cocompact discrete hyperbolic groups, *Geom. Dedicata.* **1**6 (1984), 123–148.

28. J. W. Cannon: The theory of negatively curved spaces and groups, in 'Ergodic theory, symbolic dynamics and hyperbolic spaces'(T. Bedford, M. Keane, C. Series eds.), Oxford Univ. Press, New York, 1991, 315–369.

29. A. Cayley: On the theory of groups, *Proc. London Math. Soc.***9** (1878), 126–133.

30. B. Chandler and W. Magnus: 'The history of combinatorial group theory', Springer-Verlag, New York-Heidelberg-Berlin, 1982

31. R. Charney: Artin groups of finite type are biautomatic, *Math. Ann.* **2**92 (1992), 671–683.

32. D.J. Collins and C.F. Miller III: The conjugacy problem and subgroups of finite index, *Proc. London Math. Soc.* **3**4 (1977), 535–556.

33. M. Coornaert, T. Delzant and A. Papadopoulos: 'Notes sur les groupes hyperboliques de Gromov', Springer Lect. Notes in Math. **1**441 (1990)

34. M. Dehn: 'Papers on group theory and topology', translated and introduced by John Stillwell, Springer Verlag, Berlin, Heidelberg, New York, 1987.

35. M. Dehn: Transformationen der Kurven auf zweiseitigen Flächen, *Math. Ann.* **2** (1912), 413–421.

36. M. Dehn: Über unendliche diskontinuierliche Gruppen, *Math. Ann.* **7**1 (1912), 116–144.

37. J. Douglas: Solution of the problem of Plateau, *Trans. Amer. Math. Soc.* **3**3 (1931), 263–321.

38. C. Druţu: Filling in solvable groups and in lattices in semisimple groups, Preprint, Université de Lille, 2000.

39. C. Druţu: Cones asymptotiques et invariants de quasi-isometrie pour des espaces metriques hyperboliques, *Ann. Inst. Fourier* **5**1 (2001), 81–97.

40. C. Druţu: Remplissage dans des réseaux de Q-rang 1 et dans des groupes résolubles, *Pacific J. Math.* **1**85 (1998), 269–305.

41. C. Druţu: Quasi-isometry invariants and asymptotic cones, *Internat. J. Algebra Comput.*, to appear.

42. D.B.A. Epstein, J.W. Cannon, D.F. Holt, S.V.F. Levy, M.S. Paterson, W.P. Thurston: 'Word processing in groups', Jones and Bartlett, Boston MA., 1992.

43. B. Farb: The extrinsic geometry of subgroups and the generalized word problem, *Proc. London Math. Soc.* **6**8 (1994), 577–593.

44. H. Federer: 'Geometric Measure Theory', Springer-Verlag, New York, 1969.

45. S.M. Gersten: Dehn functions and l_1-norms of finite presentations, in 'Algorithms and classification in combinatorial group theory' MSRI Publ. 23, Berkeley, CA, 1989, 195–224,.

46. S. M. Gersten: Isoperimetric and isodiametric functions of finite presentations, in "Geometric group theory", vol. 1, LMS lecture notes 181, (G. Niblo and M. Roller, eds.), Camb. Univ. Press, 1993.

47. S.M. Gersten: Preservation and distortion of area in finitely presented groups, *GAFA (Geom. Funct. Anal.)* **6** (1996), 301–345.

48. S.M. Gersten: Cohomological lower bounds for isoperimetric functions on groups, *Topology* **3**7 (1998), 1031–1072.

49. S.M. Gersten: The double exponential theorem for isodiametric and isoperimetric functions, *Internat. J. Algebra Comput.* **1** (1991), 321–327.

50. S.M. Gersten and T. Riley: Filling length in finitely presentable groups, *Geom. Dedicata.*, to appear.

51. E. Ghys and P. de la Harpe (ed): 'Sur les groupes hyperboliques d'après Mikhael Gromov', Prog. in Math. **8**3, Birkhäuser, Boston, 1990

52. R.E. Gompf: A new construction of symplectic manifolds, *Ann. of Math.* **1**42 (1995), 527–595.

53. M. Gromov: Infinite groups as geometric objects, Proc. ICM, Warsaw 1978, PWN, Warsaw, 1984, 385–392.

54. M. Gromov: 'Partial Differential Relations', Springer-Verlag, Berlin-New York, 1986.

55. M. Gromov: Hyperbolic groups, in 'Essays on group theory', MSRI Publ. No. 8 (S. M. Gersten, ed.), Springer-Verlag, New York, 1987.

56. M. Gromov: 'Asymptotic invariants of infinite groups', Geometric group theory, vol. 2, LMS lecture notes 182, (G. Niblo and M. Roller, eds.), Camb. Univ. Press, 1993.

57. F. Grunewald: On some groups which cannot be finitely presented, *J. London*

Math. Soc. **1**7 (1978), 427–436.

58. A. Hatcher and K. Vogtmann: Isoperimetric inequalities for automorphism groups of free groups, *Pacific J. Math.* **1**73 (1996), 425–441.

59. C. Hidber: Isoperimetric functions of finitely generated nilpotent groups and their amalgams, PhD Thesis 12166, ETH Zürich, 1997.

60. S. Hildebrandt: *Boundary value problems for minimal surfaces*, in 'Geometry V', Encyclopaedia Math. Sci. vol. 90, Springer, Berlin, 1997, pp. 153–237.

61. E. R. van Kampen: On some lemmas in the theory of groups, *Amer. J. Math.* **5**5 (1933), 268-273.

62. E. Leuzinger and Ch. Pittet: Isoperimetric inequalities for lattices in semisimple Lie groups of rank 2, *GAFA (Geom. Funct. Anal.)* **6** (1996), 489–511.

63. E. Leuzinger and Ch. Pittet: Quadratic Dehn functions for solvable groups, Preprint, Toulouse, 2000.

64. R.C. Lyndon: Cohomology theory of groups with a single defining relation, *Annals of Math.* **5**2, 650–665.

65. R.C. Lyndon: On Dehn's algorithm, *Math. Ann.*, **1**66 (1966), 208–228.

66. R.C. Lyndon and P.E. Schupp: 'Combinatorial Group Theory', Springer-Verlag, New York, 1977.

67. I. G. Lysenok: On some algorithmic properties of hyperbolic groups, *Math. USSR Izv.* **3**5 (1990), 145–163. (English translation of Izv. **5**3 (1989))

68. A. M. Macbeath: Groups of homeomorphisms of a simply connected space, *Ann. of Math.* **7**9 (1964), 473–488.

69. N. Macura: Quadratic isoperimetric inequality for mapping tori of polynomially growing automorphisms of free groups, GAFA (Geom. funct. anal.) **1**0 (2000), 874-901.

70. A. A. Markov: Insolubility of the problem of homeomorphy, Proc. ICM Cambridge 1958, pp. 300-306, Cambridge Univ. Press., 1960.

71. W.H. Meeks III and S-T Yau: The classical Plateau problem and the topology of three-dimensional manifolds, *Topology* **2**1 (1982), 409–442.

72. C. F. Miller III: 'On group-theoretic decision problems and their classification', Annals of Math. Studies, No. 68, Princeton University Press (1971).

73. C. F. Miller III: Decision problems for groups: survey and reflections, in 'Algorithms and Classification in Combinatorial Group Theory' (eds. G Baumslag and C. F. Miller III), MSRI Publications No. 23, Springer-Verlag (1992), 1-59.

74. J. Milnor: A note on the fundamental group, *J. Diff. Geom.* **2** (1968), 1–7.

75. C.B. Morrey: The problem of Plateau in a Riemann manifold, *Ann. of Math.* **4**9 (1948), 807–851.

76. L. Mosher: Mapping class groups are automatic, *Ann. of Math.* **1**42 (1995), 303–384.

77. W. Neumann and L. Reeves: Central extensions of word hyperbolic groups, *Ann. of Math.* **1**45 (1997), 183–192.

78. I.G. Nikolaev: Solution of the Plateau problem in spaces of curvature at most K (Russian), *Sibirsk. Mat. Zh.* **2**0 (1979), 345–353.

79. J.C.C. Nitsche: 'Lectures on minimal surfaces, Vol 1.', Camb. Univ. Press, 1989.

80. I. Niven: 'Irrational Numbers', The Carus Mathematical Monographs **1**1, MAA, 1967.

81. A. Yu Ol'shanskii: Hyperbolicity of groups with subquadratic isoperimetric inequalities, *Intl. J. Alg. Comput.* **1** (1991), 281–290.

82. A. Yu Ol'shanskii: Distortion functions for subgroups, in 'Geometric Group Theory Down Under' (Canberra, 1996), 281–291, de Gruyter, Berlin, 1999.

83. A. Yu Ol'shanskii and M. Sapir: Quadratic isoperimetric functions for Heisenberg groups: a combinatorial proof, *J. Math. Sci.* (New York) **9**3 (1999), 921–927.

84. A. Yu Ol'shanskii and M. Sapir: Length and area functions on groups and quasi-isometric Higman embeddings, *Internat. J. Algebra Comput.*, to appear.

85. D.V. Osin: Subgroup distortions in nilpotent groups, *Comm. Algebra*, to appear.

86. R. Osserman: A proof of the regularity everywhere of the classical solution to Plateau's problem, *Ann. of Math.* **9**1 (1970), 550–569.

87. P. Papasoglu: On the sub-quadratic isoperimetric inequality, in 'Geometric group theory' (R. Charney, M. Davis, M. Shapiro, eds.), de Gruyter, Berlin-New York, 1995.

88. P. Papasoglou: Asymptotic cones and the quadratic isoperimetric inequality, *J. Diff. Geom.* **4**4 (1996), 789–806.

89. Ch. Pittet: Isoperimetric inequalities in nilpotent groups, *J. London Math. Soc.* **5**5 (1997), 588–600.

90. J.A.F. Plateau: 'Statique Experimentale et Théorique des Liquides Soumis aux Seules Forces Moleculaires', Paris, Gauthier-Villars, 1873.

91. T. Rado: On Plateau's problem, *Ann. of Math.* **3**1 (1930), 457–469.

92. T. Rado: 'On the Problem of Plateau', Ergeb. der Math. Wiss. 2, Springer-Verlag, Berlin, 1953.

93. Yu. G. Reshetnyak: Non-expansive maps in a space of curvature no greater than K, *Sib. Mat. J.* **9** (1968), 683–689.

94. J.J. Rotman: 'An introduction to the theory of groups', Fourth edition, GTM 148, Springer-Verlag, New York, 1995.

95. M. Sapir, J-C. Birget, E. Rips: Dehn functions of finitely presented groups, *Ann. of Math.*, to appear.

96. Z. Sela: The isomorphism problem for hyperbolic groups I, *Ann. of Math.(2)* **1**41 (1995), 217–283.

97. J-P Serre: 'Trees', Springer-Verlag, Berlin–Heidelberg–New York, 1977.

98. H. Short (ed.): Notes on word hyperbolic groups, in 'Group theory from a geometrical viewpoint' (E. Ghys, A. Haefliger, A. Verjovsky, eds.), pp. 3-64, World Scientific, Singapore, 1991.

99. W. P. Thurston: 'The geometry and topology of 3-manifolds', Lecture notes, Princeton University, 1979.

100. W. P. Thurston: Three dimensional manifolds, Kleinian groups and hyperbolic geometry, *Bull. Amer. Math. Soc.* **6** (1982), 357-381.